USE OF DNA FINGERPRINTING AND NOVEL MOLECULAR METHODS TO IDENTIFY SOURCES OF *Escherichia coli* IN THE ENVIRONMENT

By

TROY M. SCOTT

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2002

Copyright 2002

by

Troy M. Scott

ACKNOWLEDGMENTS

I would first like to acknowledge my mentor, Dr. Samuel R. Farrah, for his guidance and friendship throughout my graduate career. He has always both encouraged and supported creativity in my research. I would also like to thank the members of my graduate committee, Dr. Edward M. Hoffmann, Dr. Phillip Achey, and Dr. Paul A. Gulig for their guidance, assistance, and willingness to lend an ear. Special thanks goes to Dr. Kenneth Portier for his statistical analyses, and to Dr. Thomas Bobik, Greg Havemann, and Stuart Underwood for lending their expertise in various aspects of molecular biology. This study would not have been possible without the insight and assistance of my friends and colleagues, Dr. George Lukasik, Dr. Salina Parveen, Andrew Koo, and Jack Shelton. I would also like to thank my family and friends, the entire Department of Microbiology and Cell Science, my fellow graduate students, Cheryl Boice, Johnny Davis, and Stephanie Sheperd, and the Engineering Research Center (ERC) for Particle Science and Technology for their help and support (technical, financial, and otherwise). Finally, I would like to acknowledge my loving girlfriend, Alexandra Maistrellis, for being understanding and supportive throughout this entire process.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS iii

LIST OF TABLES vii

LIST OF FIGURES ix

ABSTRACT x

1 INTRODUCTION 1

Disease-Causing Bacteria and their Detection in the Environment ..1
Escherichia coli 2
Shigella 3
Salmonella 5
Vibrio 6
Campylobacter 7
Disease-Causing Viruses and their Detection in the Environment ...8
Adenovirus 9
Poliovirus 9
Hepatitis A 11
Rotavirus 12
Norwalk Agent 12
Disease-Causing Protozoa and their Detection in the Environment 13
Entamoeba histolytica 13
Giardia lamblia 14
Cryptosporidium parvum 14
Microsporidia 16
Detection of Microbial Indicators of Fecal Pollution 16
Methodology and Experimental Rationale 22

2 GEOGRAPHICAL VARIATION IN RIBOTYPE PROFILES OF *Escherichia coli* ISOLATED FROM HUMANS, SWINE, POULTRY, BEEF, AND DAIRY CATTLE IN FLORIDA 25

Materials and Methods 26
Collection of Fecal Samples from Livestock and Humans 26
Isolation of *Escherichia coli* 27
Selection of *E. coli* Reference Strains 27
DNA Extraction 29
Determination of DNA Concentration 29
Restriction Enzyme Digestion 29
Southern Blot Analysis 29
Probe Preparation 30
Hybridization and Detection 30
Statistical Analysis 30
Results 31
Discussion 34

3 DEVELOPMENT OF A RAPID, PCR-BASED METHOD FOR USE IN IDENTIFYING SOURCES OF FECAL POLLUTION 40

Characteristics of *Escherichia coli* Fimbrial Adhesins 41
Experimental Approach 43
Materials and Methods 46
Collection of Fecal Samples from Livestock and Humans 46
Isolation of *Escherichia coli* 47
AFLP Primers and Adapters 47
Preparation of Genomic DNA and Adapter Ligation 48
AFLP Reactions 48
Repetitive Element Polymerase Chain Reaction (Rep-PCR) 49
PCR Amplification of rDNA Intergenic Spacer Regions 49
PCR Amplication of *FimA* and *FimH* Genes in *E. coli* 50
PCR Amplification of *papG* Genes in *E. coli* 53
Restriction Enzyme Analysis of rDNA, *FimA*, and *FimH* genes .. 53
Gel Analysis 55
Excision of Unique Bands and DNA Recovery 55
Cloning of PCR Products and Fimbrial Genes 55
Analysis of Positive Clones 56
Plasmid Isolation and DNA Sequencing 56
Sequence and Phylogenetic Analyses of *FimA* Gene 56
PCR Amplification of DNA Sequences Within the *FimA* Gene 58
Results 58
Collection of *E. coli* Isolated from Florida Livestock and Human Sources 58
AFLP Reactions 60
Screening of Primers Derived from AFLP Fragments 60
Repetitive Element Polymerase Chain Reaction (Rep-PCR) 66
PCR Amplification of rDNA Intergenic Spacer Regions 66

Cloning and sequence analysis of papG genes in *E. coli* 68
PCR Amplification of *FimA* and *FimH* in *E. coli* 68
RFLP Profiles of PCR-Amplified rDNA and *FimA* Genes 68
Cloning and Sequence Analysis of *FimH* genes in *E. coli* 69
DNA Alignment Analysis of *FimA* Genes 73
Amino Acid Analysis of *FimA* Genes 78
Hydrophilicity Profile of Type-1 Fimbrial Subunit (*FimA*) Protein 82
Phylogenetic Analysis of *FimA* Genes 82
Identification of *E. coli* Isolates by TMS and REV Primer Sets .. 86
Characterization of Plasmids Containing Portion of *FimA* Gene. 89
Discussion 90

4 SUMMARY AND CONCLUSIONS 95

REFERENCES 103

BIOGRAPHICAL SKETCH 115

LIST OF TABLES

Table — Page

1. Number and sources of *E. coli* isolates used in this study 28
2. Species-level classification of ribotype profiles generated from *E. coli* isolated from livestock 32
3. Human/nonhuman classification of ribotype profiles generated from *E. coli* isolated from livestock 33
4. Primer sets used to amplify ribosomal DNA intergenic spacer regions in *Escherichia coli* 51
5. PCR primers used to amplify the *FimA* and *FimH* genes in *E. coli* 52
6. Primer sets used to amplify variant *papG* genes in *E. coli* 54
7. *Escherichia coli* isolates used in phylogenetic analyses..... 57
8. Sequences of TMS and REV primer sets 59
9. Primer sets derived from unique DNA fragments generated by Amplified Fragment Length Polymorphism 63
10. Human/nonhuman classification of *Escherichia coli* using TMS and REV primer sets 88

LIST OF FIGURES

Figure — Page

1. Spatial plot of first two canonical dimensions for ribotype profiles of *E. coli* isolated from livestock 35

2. Spatial plot of first and third canonical dimensions for ribotype profiles of *E. coli* isolated from livestock 36

3. Spatial plot of second and third canonical dimensions for ribotype profiles of *E. coli* isolated from livestock..... 37

4. AFLP patterns observed in 2% agarose gel 61

5. AFLP patterns observed in 5% polyacrylamide gel 62

6. Rep-PCR fingerprints generated using primer set A..... 63

7. RFLP of rep-PCR products generated using primer set A 65

8. Amplification of rDNA intergenic spacer regions In *Escherichia coli*..... 67

9. Amplification of the gene encoding the major structural component (*FimA*) of type-1 fimbriae in *Escherichia coli* 70

10. Amplification of the gene encoding the lectin component (*FimH*) of type-1 fimbriae In *Escherichia coli* 71

11. RFLP patterns observed after *AluI* restriction enzyme digestion of the *FimA* gene in *Escherichia coli*..... 72

12. DNA sequence alignments of *FimA* genes from group I *E. coli* isolates 74

13. DNA sequence alignments of *FimA* genes from group II *E. coli* isolates..... 75

14. DNA sequence alignments of *FimA* genes from group III *E. coli* isolates 76

15. DNA sequence alignments of *FimA* genes from *E. coli* (Groups I, II, and III) 77

16. DNA and amino acid sequence of *FimA* gene from group I *Escherichia coli* isolates 79

17. DNA and amino acid sequence of *FimA* gene from group II *Escherichia coli* isolates 80

18. DNA and amino acid sequence of *FimA* gene from group III *Escherichia coli* isolates 81

19. Hydrophilicity profile of type-1 fimbrial subunit protein (*FimA*) in *Escherichia coli* 83

20. Unrooted phylogenetic tree showing genetic relationship between *FimA* genes from human and animal *Escherichia coli* isolates sequenced in this study 84

21. Phylogenetic tree showing genetic relationship between *FimA* genes from human and animal *Escherichia coli* isolates sequenced in this study and in published literature 85

22. Schematic representation of sequence of primer usage to identify Human and Animal-derived *FimA* genes in *Escherichia coli* 87

23. Gel electrophoresis of plasmid pFA01 (a) and PCR product pCR*FimA* (b) using TMS1 and REVgg primer set 91

24. Sequence alignment of pCR*FimA* and group I *FimA* gene 92

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

USE OF DNA FINGERPRINTING AND NOVEL MOLECULAR METHODS TO IDENTIFY SOURCES OF *ESCHERICHIA COLI* IN THE ENVIRONMENT

By

Troy M. Scott

May 2002

Chairman: Samuel R. Farrah
Major Department: Microbiology and Cell Science

Fecal pollution affects the quality and safety of many water systems and can originate from both human and non-human sources including farm runoff, wildlife impact, agricultural waste, inadequate wastewater treatment, improper waste disposal, and septic failure. Understanding the origin of fecal pollution is paramount in assessing the proper risk and remedial action necessary after the problem has been identified. Feces from humans and animals contain a variety of pathogenic microorganisms, and many of these pathogens are not readily detectable in the environment by conventional methods as they are often present in very low numbers. In addition, different pathogens are harbored by different animal species, making identifying

the type of pollution necessary for proper risk assessments to be performed. The prediction of the presence of pathogens is typically performed by the detection of established microbial indicators; however, these indicators are not adequate in identifying sources of pollution. Consequently, when they are detected in the environment using conventional tests, the source and the full extent of potential human health risks cannot be determined.

The first half of this study extended previous research using ribotyping to differentiate *E. coli* (a well-established fecal indicator) isolated from various animal species by including a greater number of isolates collected from a larger geographic region. As a result, it was determined that this method was not sufficient for differentiating sources of *E. coli* at the host species level outside of a confined watershed. For this reason and because ribotyping is time consuming and expensive, the second half of this study sought a more rapid molecular bacterial source tracking method by investigating the possibility that specific genetic differences exist between *Escherichia coli* isolated from animals and those isolated from humans. Several methods, including AFLP, PCR ribotyping, and rep-PCR were used to analyze the entire genome of *E. coli* from various sources for significant or subtle genetic differences. Single genes were also sequenced and analyzed for differences that could be useful in

discriminating *E. coli* isolates based on host origin. Specifically, sequence analyses were performed on genes that code for fimbrial adhesins in *E. coli*. Once unique sequences were found, specific PCR primers were developed that were capable of amplifying these sequences. The result was a rapid, molecular tool that can aid in differentiating *Escherichia coli* derived from human and animal sources.

CHAPTER 1
INTRODUCTION

Fecal pollution affects the quality and safety of many water systems and can originate from both human and nonhuman sources including farm runoff, wildlife impact, agricultural waste, inadequate wastewater treatment, improper waste disposal, and septic failure. Many bacterial, viral, and protozoan pathogens may be present in the intestines of humans and animals. In order to properly assess the microbiological quality and safety of any water supply, methods must be designed to detect even small numbers of these organisms or reliable indicators of their presence. This task is especially important for water systems that are used by humans for drinking, for recreation, or in the harvesting of seafood. Of the many human microbial pathogens, several can be transmitted by contaminated water. The following is a brief overview of some common organisms that pose a risk to human health and methodologies used for their detection.

Disease-Causing Bacteria and their Detection in the Environment

Members of the family Enterobacteriaceae are gram negative bacilli that are found worldwide in soil, water, and vegetation.

Members of this family are part of the normal flora that inhabit the intestines of humans and animals. Enterobacteriaceae are capable of causing a wide variety of diseases, including septicemias, gastrointestinal infections, and urinary tract infections. The primary mode of transmission for most of these bacteria is the fecal-orai route (or via contaminated food and water). Members most often causing human disease include *Escherichia*, *Salmonella*, and *Shigella*.

Escherichia coli

Eschericia coli produces various adhesins that allow it to remain attached to cells in the intestinal and urinary tract. Although it is a member of the normal commensal flora in humans, *E. coli* is capable of causing a wide spectrum of diseases when it contains virulence-enhancing plasmids or bacteriophages or when it colonizes otherwise sterile body tissues such as the urinary tract and bladder. Urinary tract infections are a common manifestation of *E.coli* infection. The infection typically originates when bacteria from the colon contaminate the urethra, and then spread into the bladder. *E. coli* is also capable of causing gastroenteritis. The strains of *E. coli* that cause gastroenteritis are divided into five groups, each with specific sites of activity and pathogenesis. The most serious of these strains is Enterohemorrhagic *E. coli* (EHEC). Many serotypes of EHEC have been found, but the O157:H7 serotype is responsible for the most disease.

The amount of bacilli needed for EHEC infection is quite low (<100 bacilli) (Nicholls et al. 2000; Murray et al. 2001). Some infections can be mild and uncomplicated; however, a possible presentation of EHEC is hemorrhagic colitis with bloody diarrhea. The severe diarrhea and hemorrhaging associated with this strain is caused by the Shiga-like toxins, which disrupt protein synthesis in human cells. Hemolytic uremic syndrome, a severe urinary tract infection characterized by acute renal failure and hemolytic anemia, is another complication of EHEC. This is a very serious complication, which can be fatal. Contaminated water, due to fecal pollution, is a major means of transmission for this organism. While most *E. coli* are considered harmless or opportunistic pathogens, their presence is most often used as an indicator of the presence of other fecal pathogens. The detection of Enterohemorrhagic *E. coli* (particularly O157:H7) is of paramount importance and both molecular (PCR, probe hybridization) and biochemical (inability to metabolize MUG or ferment sorbitol) methods have been used to specifically identify this pathogen.

Shigella

Shigella species are also transmitted by the fecal-oral route and are capable of causing severe, bloody, mucoid diarrhea with fever. Shigellosis can spread rapidly in a community with poor sanitation standards. Epidemic outbreaks often occur in crowded institutions,

such as daycare centers. The amount of bacteria required to cause an infection is low (about 300 bacilli) (Holcomb et al. 1999). The bacteria first attach and invade the M (microfold) cells in the Peyer's patches of the large intestines. Cell-to-cell transmission via the epithelial lining spreads the infection. *Shigella dysenteriae* produces Shiga toxin, an exotoxin nearly identical to that produced by Enterohemorrhagic *E. coli* and clinical symptoms are nearly identical to those caused by EHEC. While most cases of Shigellosis are caused by contaminated food or by direct person-to-person contact, the organism may also be spread by contaminated drinking water. The detection of *Shigella* spp. in environmental samples is difficult due to the low number of organisms present and the presence of a large number of background flora. Although enrichment and direct-detection techniques are available, they are difficult to perform. Additionally, the growth of coliform bacteria in environmental samples is antagonistic to the growth of *Shigella* (Blostein et al. 1991). Because the infective dose is so low, however, the presence of any amount of *Shigella* poses a threat to human health (Stutman 1994). The preferred direct method of detection of *Shigella* species in drinking water is PCR (Frankel et al. 1990). Another method of detection involves a colony blot immunoassay (Szakal et al. 2001). These

methods are more sensitive in identifying pathogen-containing samples than the traditional culture based methods.

Salmonella

Salmonella species are found in humans and a wide variety of warm-blooded and cold-blooded animals, both wild and domestic. Reservoirs are maintained by animal-animal spread. Humans obtain *Salmonella* infections from eating or drinking contaminated food or water and by direct fecal-oral spread. A large number of bacilli must be ingested in order for symptomatic disease to ensue (Stutman 1994). Enteritis is the most common form of infection caused by *Salmonella enterica* species. Symptoms include nonbloody diarrhea, nausea, and vomiting.

Salmonella typhi infection occurs after ingestion of contaminated food or water. The syndrome that follows is called typhoid fever. Symptoms include a gradually increasing fever, headache, malaise, and finally gastrointestinal symptoms. Septicemia, diarrhea, shock, and infection of the gallbladder may also occur. *Salmonella typhi* infections can be severe and fatal. The methods used in the detection of *Salmonella* in water are not standardized and provide variable degrees of reliability. Selective enrichment and identification procedures are tedious and labor-intensive. Although these tests can be performed with some degree of accuracy, a negative test result

cannot imply the absence of salmonellae or other enteric pathogens. Various molecular methods for the detection of *Salmonella* have also been proposed. Detection of *Salmonella* in the marine environment and in shellfish is preferentially done by PCR. However, although PCR is rapid and specific, marine samples often result in false-positive results (Dupray et al. 1997). As with any PCR assay, the test does not discriminate between viable and nonviable organisms (Dupray et al. 1997). Immunomagnetic separation has also been used to detect *Salmonella typhi* in food and water. This technique so far has proven to be more rapid than the traditional culture-based methods, which are slow and often ambiguous (Yu et al. 1996).

Vibrio

Vibrio cholerae is a gram negative, facultative anaerobe of the Vibrionaceae family. This bacterium grows in marine and estuarine environments worldwide but disease is usually concentrated in areas of poor sanitation and In underdeveloped countries. The primary mode of transmission for *Vibrio cholerae* is via the ingestion of contaminated food (especially shellfish) or water. A high inoculum of bacteria is required to cause infection (10^6-10^8 organisms) (Seas et al. 2000). Soon after infection with *Vibrio cholerae*, large amounts of watery diarrhea and vomiting occur. The severe fluid loss results in loss of electrolytes and dehydration. Shock and renal failure can also occur.

Standard Methods are available for the concentration and detection of *V. cholerae*. Usually, water samples are concentrated and enriched and plated on selective and differential media for presumptive identification. Verification of *V. cholerae* biotypes is usually achieved by serological analysis (Clesceri et al. 1998). DNA probes are useful in identifying which strains contain the cholera toxin gene (Clesceri et al. 1998; Hsu et al. 2001).

Campylobacter

The genus *Campylobacter* consists of gram-negative bacteria that are motile due to the presence of a polar flagellum. *Campylobacter* infections are zoonotic, with many different animals acting as reservoir hosts. Humans acquire *Campylobacter* infections through the ingestion of contaminated food and water. Fecal-oral transmission from person-to-person can also occur. Campylobacteriosis is currently the most commonly reported diarrheal illness in the United States (Stutman 1994). *Campylobacter jejuni* infections result in acute enteritis with abdominal pain, bloody diarrhea, and fever. The infection also results in a serious loss of electrolytes as well as dehydration. The method of detection of *Campylobacter* in drinking water must be performed on large samples and involves filtration/concentration, enrichment, and then detection

by use of selective enriched media containing antibiotics, biochemical tests, or by molecular methods such as PCR (Waage et al. 1999).

Disease-Causing Viruses and their Detection in the Environment

It is estimated that over 100 human enteric viruses can be transmitted by human feces (Puig et al. 1994). Most of these viruses infect the gastrointestinal tract and are transmitted via person-to-person contact or by contaminated food and water. The viruses known to be present in relatively large numbers in human feces include polioviruses, coxsackieviruses, echoviruses, adenoviruses, reoviruses, rotaviruses, Hepatitis A virus, and Norwalk-like viruses. Although these viruses are present in domestic wastewater and areas impacted by reclaimed water year-round, it has been shown that the presence of enteroviruses reach peak levels during the late summer and early fall while Norwalk-type viruses and Rotaviruses predominate during the colder winter months (Rotbart et al. 1999). Once in the environment, viruses cannot replicate and their numbers either diminish or remain constant. Even a low level of viruses in the environment can pose a risk to human health, however, as many have a very low (<10) infectious dose (Murray et al. 2001). Detection of viruses in environmental samples consists of collection and concentration of viruses from a large sample volume (usually done by filtration)

followed by a second concentration step and finally by procedures designed for specific or general identification of viable virions.

Adenovirus

Adenoviruses belong to the family Adenoviridae. The genome of the adenovirus consists of a single linear molecule of dsDNA. Adenoviruses are capable of causing numerous clinical syndromes including respiratory infections, ocular infections, genitourinary infections, and enteric infections. The serotypes responsible for the enteric infections include 31, 40 and 41. Gastroenteritis due to adenovirus is most common in children. These enteric viruses are spread by the fecal-oral route and epidemics have been recorded in institutions such as daycare centers and hospitals. Swimming pools, drinking water, and wastewater represent possible reservoirs for adenoviruses. Adenoviruses are highly resistant to disinfection and are commonly found in treated domestic wastewater (Pina et al. 1998). Detection of adenoviruses in cell culture is difficult and this method is often coupled with PCR to increase sensitivity (Cho et al. 2000; Puig et al. 1994).

Poliovirus

Polioviruses are enteroviruses in the family Picornaviridae. The poliovirus genome consists of +sense ssRNA and is contained within a naked icosahedral capsid. Enteroviruses infect the alimentary canal,

and are very resistant to the conditions in the gut. Poliovirus is spread predominantly through the fecal-oral route. Although vaccination has virtually eradicated the virus in developed nations, outbreaks have resulted from the consumption of contaminated water supplies and sewage (Knolle et al. 1995). Most poliovirus infections cause minor illness characterized by fever, vomiting, and a sore throat. However, in some cases, muscle stiffness and pain lead to the rapid development of flaccid paralysis. Death may occur due to cardiac or respiratory failure. In those that survive the disease, there is a good chance of permanent paralysis or post polio syndrome. One method of detecting poliovirus and other enteroviruses in water or sewage is by the use of absorption-elution techniques followed by cell culture detection (Clesceri et al. 1998, Kittigul et al. 2000). Enteroviruses can also be detected using direct RT-PCR, but this requires a concentrated sample and does not differentiate between viable and nonviable organisms. The problem with these methods, however, is that they involve several time-consuming steps (Legeay et al. 2000). Another problem with direct RT-PCR (or conventional cell culture) is that these techniques may lead to erroneous results in environmental samples (Reynolds et al. 2001). An integrated cell culture/polymerase chain reaction methodology (ICC/PCR) is very useful in detecting the

presence of viable enteroviruses with greater accuracy (Reynolds et al. 2001).

Hepatitis A

Hepatitis A belongs to the genus Hepatovirus and is a member of the Picornaviridae family. Hepatitis A is spread through the fecal-oral route and is found only in the human reservoir. The diseases caused by this virus are endemic in areas of overcrowding and poor sanitation (such as developing countries). Major common-source outbreaks have occurred due to contaminated wells and water supplies. Symptoms of this infection include fever, malaise, nausea, dark urine, and abdominal pain. Jaundice occurs as well, representing infection of the liver. In a small number of cases, death may occur due to fulminant hepatitis, but more often, all nonlethal infections resolve with complete regeneration of the damaged liver parenchyma (Rose et al. 2000). A killed HAV vaccine has been approved by the U.S. Food and Drug Administration and is available for high-risk individuals, such as those traveling to endemic regions (Murray et al. 2001). The methods of detection of Hepatitis A are the similar to those used for poliovirus, but are more involved. RT-PCR is commonly utilized (Kingsley and Richards, 2001; Kittigul et al. 2000).

Rotavirus

Rotaviruses are members of the family Reoviridae. They are non-enveloped and their genomes contain 10-12 molecules of dsRNA. During an active infection, rotaviruses are shed in extremely large numbers in feces. Transmission of rotavirus occurs mainly through contact with feces. Waterborne epidemics do occur, and institutions such as daycare centers and hospitals have high numbers of infections. Infection with rotavirus can be asymptomatic or symptomatic. Symptoms of rotavirus infection include vomiting, then diarrhea with severe dehydration. Death is rare in well-developed countries, but significantly higher in underdeveloped countries (Katyal et al. 2000). Rotaviruses can be detected in drinking water by filtration and concentration followed by reverse transcription PCR analysis (Gratacap-Cavallier et al. 2000). A modified adsorption-elution technique, followed by enzyme-linked immunosorbent assay (ELISA) for concentration of rotavirus in sewage and drinking water is also used (Kittigul et al. 2000).

Norwalk Agent

The Norwalk virus is a member of the family Caliciviridae. The Norwalk virus is non-enveloped and the genome consists of +ssRNA. The disease caused by the Norwalk virus has a very short incubation period and the infection itself is short-lived. Symptoms include

nausea, vomiting, diarrhea, low fever, and abdominal cramps. Common-source outbreaks frequently happen via fecal contamination of food and water. Major outbreaks have occurred due to the consumption of raw shellfish taken from sewage-polluted estuaries (Le Guyader et al. 2000). Also, outbreaks have been attributed to discharge of sewers into drinking supplies. Norwalk virus can also be detected in environmental samples by filtration and concentration followed by reverse transcription-PCR (Atmar and Estes, 2001) as well as RT-PCR-oligoprobe amplification (Schwab et al. 2001).

Disease-Causing Protozoa and their Detection in the Environment

Entamoeba histolytica

Entamoeba histolytica infections occur worldwide and are the third leading cause of death among parasitic infections. The cyst form of this organism is infective to humans and the cysts have the ability to survive in food and water. Transmission in water is common in developing countries, where much of the water supply is untreated and contaminated with feces. The use of human feces for fertilizer is also an important source of infection. Several clinical presentations of *E. histolytica* are known. Infection may be asymptomatic in some individuals and fatal in others. Acute amebic colitis involving frequent bloody stools and fever may develop. A similar syndrome, fulminant colitis, may also occur. Ameboma, characterized by asymptomatic

lesions is yet another presentation. If left untreated, amoebae can disseminate throughout the body causing disease in the brain, liver, heart, and lungs (Cook 1997; Natarajan et al. 2000). Detection of *Entamoeba histolytica* can be performed by direct microscopic examination or by various molecular assays, including ELISA and PCR (Evangelopoulos et al. 2001; Schunk et al. 2001; Zindrou et al. 2001).

Giardia lamblia

Giardia lamblia is the most commonly isolated intestinal parasite in the world and it is especially prevalent in children in underdeveloped countries (Gardner et al. 2001). Outbreaks of *Giardia* are associated with the ingestion of unfiltered, inadequately chlorinated water. Waterborne transmission causes giardiasis in travelers in endemic countries. Symptoms of *Giardia lamblia* infection begin with the onset of intestinal uneasiness, followed by nausea. Explosive, watery, foul-smelling diarrhea follows. Abdominal cramps, low-grade fever, and chills may also occur in some cases.

Cryptosporidium parvum

Cryptosporidium parvum is an intestinal parasite found worldwide. Humans obtain *Cryptosporidium* infections upon the ingestion of oocysts present in food or water. The oocysts are widely distributed in both sewage and drinking water. Swimming pool and lake outbreaks have been identified, and as with other intestinal

parasites, travelers to countries with high rates of endemicity are vulnerable to infection. The most common symptom of *Cryptosporidium parvum* infection is diarrhea, which lasts for about one week. Other symptoms include fever, abdominal pain, and nausea. The American Society for Testing and Materials (ASTM) analytic procedure is currently one method of choice for detecting *Giardia* and *Cryptosporidium* species in source waters (Marshall et al. 1997). The method is time-consuming, difficult, and technically complex. Variables such as weather conditions may change results. The procedure involves sampling and filtration of a large sample of water, followed by concentration and flotation. The recovered particles are stained with fluorescent antibodies and observed with an ultraviolet epiflourescence microscope. Characteristic staining patterns identify *Giardia* and *Cryptosporidium*. EPA method 1623 and the Information Collection Rule (ICR) Protozoan Method are also used for recovery, identification, and quantification of *Giardia* and *Cryptosporidium* in the environment (USEPA, 1995; USEPA, 1999). Less technical, more accurate, and less time-consuming detection methods are greatly needed. Cell-culture techniques have been developed and are often supplemented by immunological or molecular verification (DiGiovanni et al. 1999; Lowery et al. 2000; Rochelle et al. 1999; Rochelle et al. 1997; Rochelle et al. 1996).

Microsporidia

Members of the Order Microsporidia have the ability to infect a wide range of vertebrate and nonvertebrate hosts. They are obligate intracellular protozoa that form highly specialized, environmentally resistant spores. Humans acquire Microsporidia infection via the ingestion of spores. Surface water is the primary environmental residence for Microsporidia. The exact route of infection is still unknown; however, the fecal-oral route has been implicated. The most common symptoms of Microsporidia infection are diarrhea, dehydration, and weight loss. Most infections are found in association with human immunodeficiency virus (HIV), implicating that an immunocompromised host facilitates infection. Restriction fragment length polymorphism analysis of PCR amplicons can be used to identify various Microsporidia (Curry, 1998). In addition, direct microscopic evaluation of specimens using immunofluorescence or electron microscopy as well as strain-specific PCR and DNA hybridization assays are emerging as effective means of identifying various species of Microsporidia (DaSilva et al. 1997; Fournier et al. 2000; Franzen and Muller, 1999; Moura et al. 1999).

Detection of Microbial Indicators of Fecal Pollution

Given the above information, and the various sources from which these pathogens can originate, it is easy to conceive why

understanding the origin of fecal pollution is necessary to properly assess the extent of risk and remedial action necessary after the problem has been identified. As stated previously, human fecal waste can contain various enteric pathogens, some of which are unique to humans, including *Salmonella typhi*, *Shigella* spp., Hepatitis A, and Norwalk-group viruses. However, other fecal pathogens are shared with animals (e.g., various serotypes of *Salmonella* and *E. coli*). Many of these fecal human pathogens are not readily detectable in the environment by conventional methods as they are often present in very low numbers. This is further complicated by the fact that many of them have a considerably low infectious dose, which renders even a low prevalence in polluted waters hazardous to human health. Therefore, the prediction of their presence is typically performed by the detection of established indicators of fecal pollution.

Escherichia coli has long been used as an indicator of fecal pollution (Geldreich, 1966). It has good characteristics of an indicator, such as not normally being pathogenic to humans, and is present at concentrations much higher than the pathogens it predicts. However, it is well established that *E. coli* is not limited to humans, but also exists in the intestines of many other warm-blooded animals (Orskov and Orskov, 1981). Consequently, when it is detected in water with conventional bacteriological tests, its source and the full extent of

potential human health risks cannot be determined. This could be remedied by the development of better testing methods and analysis techniques that can define specific sources of *E. coli*.

To meet the challenge of identifying sources of fecal pollution, various methods have been proposed. Initially, the ratio of fecal coliforms to fecal streptococci was proposed, where a ratio of >4.0 would indicate human pollution and a ratio of $\leq$ 0.7 would indicate nonhuman pollution (Geldreich and Kenner, 1969). However, this approach has proven to be unreliable due to variable survival rates of fecal streptococci species and variable sensitivity to water treatments (Pourcher et al. 1991).

Investigators have also reported that animal and human feces contain different serotypes of RNA coliphages (Furuse et al. 1981), suggesting that phages could be used to predict sources of pollution. However, its usefulness is limited, because only a small percentage of human fecal samples contain phages (Gerba 1987).

Numerous phenotypic and genotypic methods for discriminating bacteria have been suggested. These include biochemical tests (Olsen et al. 1992), phage susceptibility (Zierdt et al. 1980), outer membrane protein profiles (Barenkamp 1981), antibody reactivity (Wachsmuth 1986), fimbriation (Latham and Stamm, 1984), bacteriocin production and susceptibility, and other methods.

However, these systems have serious disadvantages, including unstable phenotypes, low sensitivity at the intraspecies level, and limited specificity.

Human-specific chemical substances have also been used in fecal source tracking. These include caffeine (Burkhardt et al. 1999), fecal sterols (Edwards et al. 1998), and laundry optical brighteners. However, the methodologies used for their detection are extremely tedious and lack the desired sensitivity needed due to the dilution of the chemical in the environment.

Multiple antibiotic resistance (MAR) is a method that has been used to differentiate *E. coli* from different sources using antibiotics commonly associated with human and animal therapy, as well as animal feed (Cooke 1976; Harwood 2000; Parveen et al. 1997; Wiggins 1996; Wiggins et al. 1999). However, ultimately, this is likely not the technique of choice, since antibiotic resistance is often carried on plasmids, which can be readily lost from cells via cultivation and storage or by changes in environmental conditions. In addition, strains from different locations may show variations in specific sensitivities due to variable antibiotic use among humans and livestock species. Furthermore, antibiotic sensitivity would not be useful in situations where the isolates under study show no significant resistance patterns, yet come from different animal species.

Pulsed field gel electrophoreses (PFGE) is a method of DNA fingerprinting. The fingerprints are generated after treatment of genomic bacterial DNA with rare-cutting restriction endonucleases. PFGE has been a very useful technique in determining bacterial relatedness but has thus far been ineffective in bacterial source tracking studies (Parveen et al. 2001).

PCR amplification of repetitive DNA sequences (rep-PCR) in *E. coli* has also been reported as a feasible method for differentiating *E. coli* from human and animal sources (Dombek et al. 2000). This method produces a reliable and reproducible fingerprint and is relatively easy to perform. However, this method also requires that a reference database be established and additional known isolates must be fingerprinted from a large geographic region in order to assess the potential universal application of this procedure.

Ribotyping is a method of DNA fingerprinting whereby highly conserved rRNA genes are identified after treatment of genomic DNA with restriction endonucleases. Ribotyping has proven to be a very useful epidemiological technique for use with various bacteria, including *E. coli* (Stull et al. 1988), *Salmonella enterica* (Olsen et al. 1992), *V. cholerae* O1 (Popovic et al. 1993), and *V. vulnificus* (Anzar et al. 1993; Tamplin et al. 1996). Ribotyping can also effectively track human and nonhuman sources of pollution with a high degree of

confidence (Carson et al. 2001; Parveen et al. 1999). This method is often used in conjunction with other methods such as multiple antibiotic resistance profiling and serological analyses to confirm bacterial origin. The ribotyping method is a two-week procedure and involves bacteriological culture and identification, DNA extraction, gel electrophoresis, ribotyping, and discriminant analysis of the DNA fragments. The success of this procedure depends on the size of the reference database to which a ribotype profile from an unknown isolate must be compared. The inability of many labs to compile a useful database is one limitation of this procedure. Although this method has proven successful, it is also expensive and time consuming.

PCR ribotyping uses PCR to amplify a fragment of the genomic DNA which includes the spacer region between the genes coding for 16S and 23S rRNA. In previous studies, sequence polymorphisms in these spacer regions have been used in differentiating closely related members of a number of bacterial genera (Cartwright et al. 1995; Kostman et al. 1992; Matar et al. 1993). *Escherichia coli* has been shown to possess seven alleles of the rRNA gene cluster (Kiss et al. 1977). Successful bacterial source-tracking using this procedure would most likely rely on the creation of a riboPCR profile database to which unknown banding patterns could be compared followed by discriminant analysis.

Rapid tests that discriminate human fecal pollution from bovine fecal pollution are currently available using Length-heterogeneity PCR (LH-PCR) and Terminal restriction fragment length polymorphism (T-RFLP) analysis to characterize members of the *Bacteriodes-Prevotella* group and the genus *Bifidobacterium* (Bernhard and Field, 2000). This study produced reliable results; however, it analyzed a limited number of isolates from only two hosts (humans and cows) from a confined geographic area.

Methodology and Experimental Rationale

The first part of this study expanded on previous research using ribotyping and discriminant analysis by including isolates collected from four species of animal (beef cattle, dairy cattle, swine, poultry) throughout Northern, Central, and Southern Florida. While published studies have reported that ribotyping is a useful tool in identifying *E. coli* based on host source, these studies used a limited number of isolates from a confined geographic region (Carson et al. 2001; Parveen et al. 1999). The present study attempted to determine whether ribotype profiles are unique to isolates from a particular host. In addition, it was investigated whether these profiles are useful outside of a confined geographic region or whether this method only provides discriminating power within a specific watershed.

The apparent success of the ribotyping procedure, as well as other genetic fingerprinting methods, has provided empirical evidence that unique DNA sequences exist within the *E. coli* genome that directly or indirectly correlate with source. In addition to ribotyping analysis, the second half of the present study examined the possibility that groups of DNA sequences exist within the genome of *E. coli* that are capable of discriminating source. Various methods of detecting differences within the bacterial genome were compared and included "brute-force" methodologies such as amplified fragment length polymorphism (AFLP) analysis, PCR-ribotyping, repetitive element polymerase chain reaction (rep-PCR), and restriction fragment length polymorphism (RFLP) analysis as well as more direct, specific methodologies such as sequence analysis of genes coding for fimbrial and nonfimbrial adhesins present on the bacterial surface that play a direct role in bacterial attachment and colonization.

The basis for the latter research approach was the possibility that host specificity is a result of differential affinity of *E. coli* for unique, specific receptors within a particular host. The gene that codes for the major structural component of type 1 fimbriae (*fimA*) and the lectin component of Type 1 fimbriae (*fimH*), as well as the genes that code for the proteins constituting the distal portion of the P1 pilus (lectin) in pyelonephritic *Escherichia coli* (papG, Classes I, II,

III) were considered as genes within the *E. coli* chromosome that may show subtle genetic variations between isolates from different sources due to differences in binding sites and receptors on different host cell surfaces.

CHAPTER 2
GEOGRAPHICAL VARIATION IN RIBOTYPE PROFILES OF *Escherichia coli* ISOLATED FROM HUMANS, SWINE, POULTRY, BEEF, AND DAIRY CATTLE IN FLORIDA

Understanding the source(s) of fecal pollution is paramount in assessing the potential risks of the contamination to human health and in properly determining the remedial actions necessary to correct the problem. While many methods have been proposed, genotypic methods provide the most reliable results, as phenotypic tests are often less stable and more sensitive to environmental factors.

Ribotyping has been used by several researchers to discriminate between closely related strains of bacteria as well as in bacterial source tracking (Carson et al. 2001; Parveen et al. 1999; Svec et al. 2001). While these studies have shown genotypic differences between human and animal-derived indicators, they have focused on isolates collected from a confined geographic area and have not addressed the question as to whether these profiles are watershed-specific or if they can be applied universally to organisms from other geographic locations.

In this study, *Escherichia coli* isolated from humans, beef cattle, dairy cattle, swine, and poultry were collected from locations in

Northern, Central, and Southern Florida and ribotyped. The intent was to determine if ribotype profiles: 1.) Were capable of discriminating the source of *E. coli* at the species level, and 2.) Were specific for a particular animal source in a specific, confined, or broad geographical region.

Materials and Methods

Collection of Fecal Samples from Livestock and Humans

Compost feces were collected from swine, poultry, dairy and cattle farms in three separate geographical regions of Florida over seasonal time intervals. Samples from dairy cattle farms were collected from retention ponds containing stall flush water located in Greenville, FL (North), Hague, FL, (Central), and Okeechobee, FL (South). Samples from beef cattle farms were collected from composite manure pits and flush water retention ponds in Lake City, FL (North), Alachua, FL (Central), and Okeechobee, FL (South). Samples from swine farms were collected from retention ponds located in Grand Ridge, FL (North), Gainesville, FL (Central), and Dade City, FL (South). Samples from chicken farms were collected from retention ponds located in Bushnell, FL (North), Dade City, FL (Central), and Zolfo Springs (South). Water samples were collected from at least three locations within the retention ponds and at least three separate samples from composite manure pits were collected from each farm.

After collection, all samples were stored at 4^0C, transported to the laboratory in refrigerated (4^0C) coolers, and processed within 24 hours. Human isolates were obtained directly from human volunteers, residential septic systems, and from sewage lines that have no animal impact. A summary of the types of isolates and samples taken is shown in Table 1.

Isolation of *Escherichia coli*

Fecal samples were streaked on MacConkey agar plates (Difco) within 24 hours of collection. Lactose-positive colonies were picked and subcultured into Tryptic Soy Broth (TSB, Difco) containing 4-Methylunbelliferyl B-D-Glucuronide (MUG) substrate (Sigma). MUG-positive isolates were presumed to be *E. coli* and were verified using the IMViC series of tests (Indole, Methyl Red, Voges-Proskauer, Citrate). Isolates exhibiting ++-- IMViC profiles were confirmed as *E. coli*.

Selection of *E. coli* Reference Strains

Several well-characterized human and nonhuman derived *Escherichia coli* from our extensive collection were selected and used in the establishment of an original database for isolate classification. These reference strains are valuable for verifying sources of *E. coli* as being either human or animal-derived. All isolates were maintained in liquid nitrogen.

Table 1. Number and sources of *E. coli* isolates used in this study

Source	Number of Isolates	Location	Sample Type
Human	84	Central	Septic tanks, fecal samples
Beef	85	Northern, Central, Southern	Lagoon, manure pit (South)
Dairy	82	Northern, Central, Southern	Lagoon
Swine	80	Northern, Central, Southern	Lagoon
Poultry	70	Northern, Central, Southern	Lagoon

DNA Extraction

E. coli isolates were grown overnight in Tryptic Soy Broth (TSB) and DNA was extracted using the Easy DNA kit (Invitrogen, Carlsbad, CA) according to manufacturer's instructions.

Determination of DNA Concentration

DNA concentration was determined using a TKO 100 fluorometer according to manufacturer's instructions.

Restriction Enzyme Digestion

Approximately 1 μg of DNA was digested with *HindIII* restriction enzyme (Roche Molecular Biochemicals) according to manufacturer's instructions. Digested DNA was separated on a 1.0% agarose gel at 30 V for 16 hours in 0.5X Tris-Borate-EDTA (TBE) buffer, stained with ethidium bromide and viewed under UV light.

Southern Blot Analysis

After electrophoresis of restriction-digested DNA, agarose gels containing restricted DNA were depurinated in 0.2M HCl for 10 minutes, denatured in 0.5M NaOH/1.5 M NaCl for 35 minutes, and neutralized in 0.5 M Trls-HCl (pH 7.2)/1.5 M NaCl (0.0001 M) disodium EDTA for 45 minutes. DNA was blotted from gels onto nylon membranes (BioRad) using a vacuum blotting system (VacuGene XL) and fixed with shortwave UV light for 5 minutes.

Probe Preparation

E. coli 16S and 23S rRNA was reverse transcribed into cDNA with avian reverse transcriptase and labeled with digoxigenin–dUTP according to the manufacturers instructions (Roche Molecular Diagnostics, Mannheim, Germany).

Hybridization and Detection

Membranes were prehybridized at 65 °C for 30 minutes in 20mM Na_2HPO_4 and 7% SDS (pH7.2) and then hybridized in the same solution containing the digoxigenin-labeled probe at 65°C for 16 hours. After hybridization, membranes were washed twice for 60 min. each time with 20mM Na_2HPO_4 and 5% SDS (pH 7.2) at 65°C followed by 2 washes for 30 min with 20mM Na_2HPO_4 and 1% SDS (pH 7.2) at 65°C. Membranes were then reacted with alkaline phosphatase conjugated anti-DIG antibody and visualized by using nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl-phosphate for colorimetric detection according to the manufacturer's instructions (Roche Molecular Diagnostics).

Statistical Analysis

RT banding profiles were read manually and DNA banding patterns were translated into binary code. Binary codes were examined using SAS (SAS Institute, Inc., Cary, N.C.) statistical discrimination methodology. Results of the discriminiation model were

summarized by use of the average rate of correct classification (ARCC) and the percentage of correctly and misclassified isolates from the classification table.

Results

Over 1800 *E. coli* were isolated and a total of 317 *E. coli* isolates were ribotyped from dairy cattle, beef cattle, swine, and poultry from northern, central, and southern Florida during the spring, summer, fall, and winter seasons. Thirty-four unique RT bands were observed and were used in the discriminate analysis. The ribotype classifications of *E. coli* isolated from the four animal types are shown in Table 2. As can be seen in the Table, the beef and dairy isolates were collectively classified as dairy and nearly half of the poultry and swine isolates were also classified as dairy. When the four animal groups were combined, however, and tested against the human/nonhuman database as a whole, 78.6% (n=249) were classified as nonhuman and the remaining 21.4% (n=68) were misclassified as human (Table 3).

Table 2. Species-level classification of ribotype profiles generated from *E. coli* isolated from livestock.

	No. of isolates (%) classified as:			
Source (# of isolates)	Beef	Dairy	Poultry	Swine
Beef (85)	5 (6)	61 (72)	18 (21)	1 (1)
Dairy (82)	4 (5)	66 (80)	8 (10)	4 (5)
Poultry (80)	5 (6)	45 (56)	26 (33)	4 (5)
Swine (70)	3 (4)	33 (47)	21 (30)	13 (19)

Table 3. Human/nonhuman classification of ribotype profiles generated from *E. coli* isolated from livestock

Source (# of isolates)	Nonhuman	Human
Beef (85)	72 (84.7)	13 (15.3)
Dairy (82)	66 (80.5)	16 (19.5)
Poultry (80)	57 (71.3)	23 (28.5)
Swine (70)	54 (77.1)	16 (22.9)
Total	249 (78.6)	68 (21.4)

In canonical discrimination, the sets of linear functions that best separate the classes in the directions of most variability were computed. If the classes are well separated, scatter plots of the canonical scores should show well-separated scatters for the classes. Figures 1, 2 and 3 are the scatter plots of the canonical scores for these data. The four animal classes are not visually separated in any of the plots. A useful scatter plot would show defined separation between the data points, with little or no overlap. This suggests that linear functions of the RT bands do not work well as discriminators. In fact, several statistical parameters were used and no set of statistical analyses was able to separate the data obtained from the different animal sources due to significant overlap of RT profiles.

Discussion

The results of this study indicate, contrary to previously published results, that the ribotyping procedure may not be useful in identifying the specific animal sources of *Escherichia coli* collected from a broad geographic region. In the present study, *E. coli* were collected from Southern, Central, and Northern Florida from beef, dairy, poultry, and swine farms. Ribotype profiles were generated from each animal in each geographic location until no profile variation was observed. These profiles were then cross-referenced within and between animal sources and assessments were made as to whether they provided

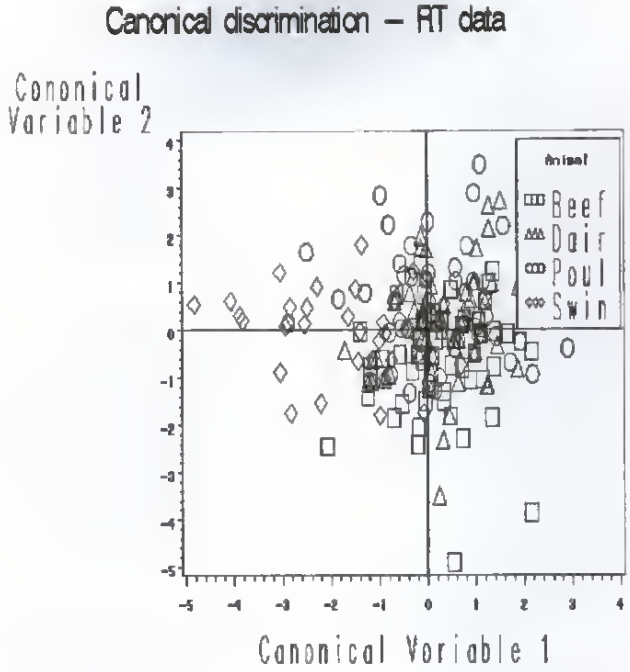

Figure 1. Spatial plot of first two canonical dimensions for ribotype profiles of *E. coli* isolated from livestock.

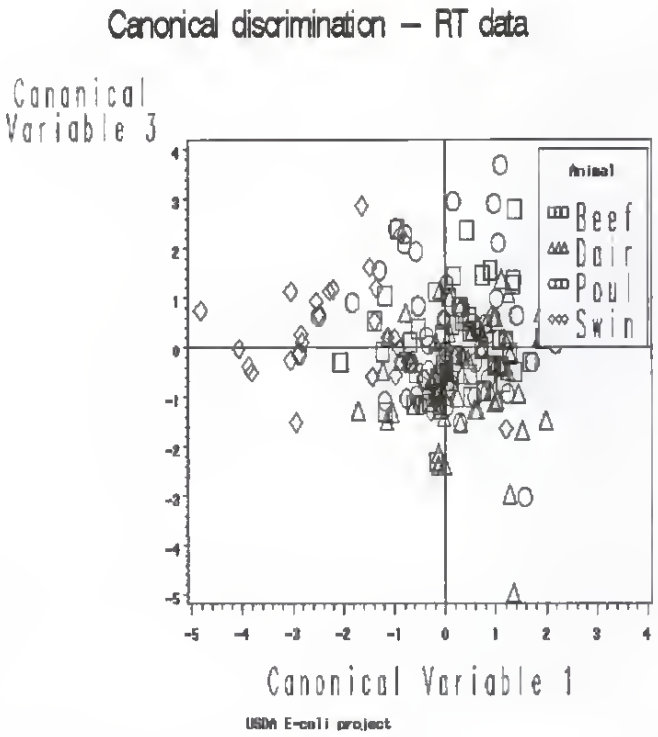

Figure 2. Spatial plot of first and third canonical dimensions for ribotype profiles of *E. coli* isolated from livestock

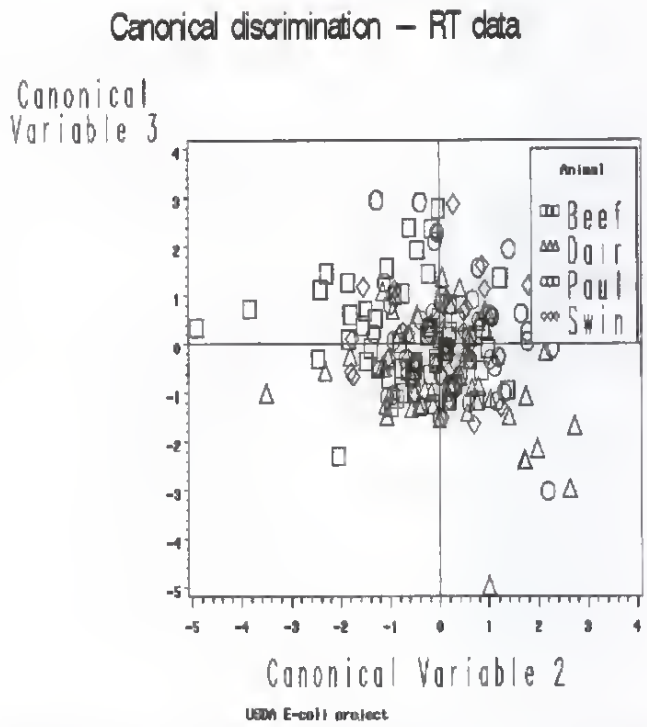

Figure 3. Spatial plot of second and third canonical dimensions for ribotype profiles of *E. coli* isolated from livestock.

discriminatory information. Overlap of ribotype profiles within and between animal groups was significant. Therefore, no single profile or group of profiles could be attributed to any particular animal source. Reasons as to why this significant overlap in RT profiles was observed which subsequently resulted in an inability to differentiate sources of *E. coli* using this procedure are not known. One significant difference between this study and a previous study by Carson et al. (2001) is the diversity of the sample collection and, in particular, the type of samples collected. Whereas Carson et al. collected individual fecal samples from a confined geographic region; we collected *E. coli* from a large geographic region; furthermore, the samples collected were environmental water samples from lagoons or compost pits. It is possible that the latter type of sample contains isolates that have been subjected to various environmental stresses, which could potentially cause mutations that adversely affect a ribotype profile. Therefore, a combination of geographic and environmental variation may play a significant role in affecting the ability of ribotyping to identify sources of *Escherichia coli* in the environment.

One significant result of this study was the fact that ribotype profiles from *E. coli* isolated from animals still differed significantly from those obtained from human isolates. Therefore, it appears that the method may have far reaching capacity for discriminating strictly

between isolates derived from animals from those derived from humans. Overall, the correct classification of *E. coli* as being either human or animal-derived was greater than 78%. Although there is not an established standard of accuracy that has been defined for any bacterial source tracking method, any method with a confidence level over 50% has been considered as a worthwhile tool for predicting the potential sources of fecal pollution in environmental waters. Therefore, ribotyping continues to have merit as a viable molecular tool to be used for this purpose.

CHAPTER 3
DEVELOPMENT OF A RAPID, PCR-BASED METHOD FOR USE IN IDENTIFYING SOURCES OF FECAL POLLUTION

Many of the molecular methods that are currently used to identify sources of fecal pollution are time-consuming, or involve the use of expensive or specialized equipment. A rapid test method would provide water quality managers with nearly real-time information so that measures to correct the contamination problem could be taken while the problem still exists or is easy to identify. Polymerase Chain Reaction (PCR)- based methodologies are fast and reliable, and require only specific primers, and a thermocycler, which is standard equipment in most basic microbiology laboratories.

The purpose of this study was to develop a rapid method capable of identifying specific regions within the genome of *Escherichia coli* that can be used to differentiate organisms isolated from either human or animal sources. Several methods, including AFLP, PCR ribotyping, and rep-PCR were used to analyze the entire genome of *E. coli* for significant or subtle genetic differences between organisms isolated from different animal sources. Single genes were also sequenced and analyzed for differences that could be useful in discriminating *E. coli*

isolates based on host origin. Specifically, sequence analyses were performed on genes that code for fimbrial adhesins in *E. coli*. Once unique sequences were found, specific PCR primers were developed that are capable of identifying sources of *E. coli* and fecal pollution with a moderate to high degree of accuracy as being from either human or animal origin.

Characteristics of *Eschericia coli* Fimbrial Adhesins

The ability to adhere to mucosal surfaces on host cells is often a limiting factor involved in the pathogenicity of an organism (Savage and Fletcher 1985). The initial interaction between the mucosa of the host cell and a bacterium is likely a random event, but subsequent interactions may be the result of specific interaction between receptors on the host cell and various adhesins present on the surface of the bacterium (Savage and Fletcher 1985). Nevertheless, it is possible and likely that bacterial adhesins and fimbriae play a specific role in determining host-specificity of both pathogenic and non-pathogenic bacteria.

Type 1 fimbriae are extracellular appendages present on the surface of most strains of *Escherichia coli* and on many other members of the family Enterobacteriaceae. They are helical, mannose-specific, proteinaceous structures composed mainly of a single protein monomer (*FimA*), but also consist of minor amounts of *FimG*, *FimF*,

and *FimH* (Savage and Fletcher 1985). The assembly of the repeating major subunit (*FimA*) has a right-handed configuration with 3 1/8 subunits per revolution and a subunit pitch of 23.2 A, but the length of each individual appendage is variable and dependent upon the organism itself as well as on growth conditions (Brinton 1965).

The genetic sequence of *FimA* has been shown in previous studies to be highly polymorphic (Li et al. 1997; Peek et al. 2001). Sequence analysis of the *fimA* gene has already proven successful in identifying DNA sequences capable of identifying or differentiating Enterohemorrhagic (EHEC) *Escherichia coli* O157:H7 from other *E. coli* strains (Li et al. 1997; Roe et al. 2001). *FimA* has also been studied for its ability to differentiate between closely related strains of bacteria, including *Escherichia coli* and *Salmonella typhimurium* (Boyd and Hartl 1998). In addition, *FimA* has been the subject of recent publications that attempt to characterize and identify the source(s) and reasons for the observed genetic variability within the gene (Boyd and Hartl 1998; Peek et al. 2001).

Approximately 30-70% of *E. coli* phenotypically display Type I fimbriae (Orskov and Orskov, 1990; Tulles et al. 1992), but the gene may be present in a higher percentage of isolates as it is known that the gene is under a switch regulation by which expression can be

turned off in response to environmental and physiological conditions (Olsen et al. 1998; Schembri et al. 1998).

Another type of fimbriae, P fimbriae, are a common genotypic and phenotypic feature of uropathogenic *E. coli*. Their gene structure is similar to that of type-1 fimbriae and consists of a major protein subunit (*PapA*) as well as a fimbrial lectin (*papG*). Three different *papG* (P fimbrial lectin) alleles are present in *E. coli.* These adhesins bind specifically to Gal alpha 1-4 Gal-containing glycolipids on host cells and their specificity for variations in different receptors has been shown to correlate with host tropism (Bertin et al. 2000; Hansson et al. 1995; Haslam et al. 1994; Marklund et al. 1992;). Just as in *FimA*, the *PapA* gene is highly polymorphic and studies have investigated the reason(s) for this observed diversity (Boyd and Hartl 1998).

Experimental Approach

In this study, Amplified Fragment Length Polymorphism (AFLP), PCR-ribotyping, repetitive element polymerase chain reaction (Rep-PCR), and Restriction Fragment Length Polymorphism (RFLP) analyses were used as "brute force" methodologies designed to detect genetic differences in *Escherichia coli* isolated from human and animal sources.

The AFLP technique is based on the detection of genomic restriction fragments by PCR amplification, and can be used for DNA of any origin or complexity. Fingerprints are produced without prior

sequence knowledge using specific adapters ligated to the ends of restriction fragments. Amplification is then carried out using adapter-specific primers and fingerprints can be analyzed for specific, unique differences evidenced by the presence or absence of a specific band or banding pattern. Unlike many similar procedures, this procedure is highly reproducible and the number of fragments detected in a single reaction can be tuned by using primer sets of varying selectivity and adjusting the stringency of the PCR reaction conditions.

PCR ribotyping uses PCR to amplify a fragment of the genomic DNA which includes the spacer region between the genes coding for 16S and 23S rRNA. In previous studies, sequence length polymorphisms in these spacer regions have been used in differentiating closely related members of a number of bacterial genera (Cartwright et al. 1995; Kostman et al. 1992; Matar et al. 1993). *Escherichia coli* has been shown to possess seven alleles of the rRNA gene cluster (Kiss et al. 1977).

Repetitive element-PCR uses primers corresponding to interspersed repetitive DNA elements present in various locations within the prokaryotic genome and PCR to generate highly specific and reproducible genomic fingerprints. Three methods of repetitive sequence analysis have been used with each targeting a specific family of repetitive element. These include repetitive extragenic palindromic

(REP) sequences, Enterobacterial Repetitive Intergenic Consensus (ERIC) sequences, and BOX elements, which are extragenic repeating elements first described by Versalovic et al. (1994). The corresponding protocols are referred to as REP-PCR, ERIC-PCR, and BOX-PCR respectively, and rep-PCR collectively. Generally, the BOX primer set is used in cases where a detailed characterization is needed as this primer generates robust fingerprints, and generally yields a highly complex pattern of amplified fragments. The REP primer set generally generates a lower level of complexity, while the ERIC primer set is more sensitive to sub-optimal PCR conditions, such as the presence of contaminants in the DNA preparation. For these reasons, Bacterial Source Tracking (BST) research has initially focused on the use of the BOX primer in performing rep-PCR. The resulting genetic fingerprint using BOX-PCR contains several bands, which can subsequently be analyzed, categorized by host source, and used to construct a database to which fingerprints from unknown isolates can be compared. This method has been used previously in bacterial source tracking studies as well as in several studies designed to differentiate between closely related strains of bacteria (Dombek et al. 2000; Versalovic et al. 1994; Versalovic et al. 1998).

In a more specific approach, genes encoding fimbrial and nonfimbrial adhesins present on the surface of many strains of

Escherichia coli were amplified using PCR, purified, cloned, sequenced, and analyzed for the presence of genetic differences between organisms isolated from human and animal sources.

Several genetic fingerprints were generated using the aforementioned "brute-force" techniques, and specific, unique DNA fragments were identified using the AFLP technique. These fragments were cloned and sequenced, and PCR primers were developed that selectively amplified these sequences. In addition, regions within the genes coding for bacterial fimbriae were also used as templates for PCR primer production. These tentative probes were then screened against several, well-characterized *E. coli* from both human and animal sources. The end result was the construction of several sets of PCR primers that selectively amplify distinct portions of the *E. coli* genome and are capable of discriminating human from nonhuman-derived organisms depending on selective amplification of these gene fragments.

Materials and Methods

Collection of Fecal Samples from Livestock and Humans

Compost fecal samples were collected from dairy, beef, swine, and poultry farms located in southern, central, and northern Florida. Samples were collected quarterly with sampling dates corresponding to different seasons. Human isolates were obtained directly from human

volunteers, residential septic systems, and from sewage lines that have no animal impact.

Isolation of *Escherichia coli*

Fecal samples were streaked on MacConkey agar plates (Difco) within 24 hours of collection. Lactose-positive colonies were picked and subcultured into Tryptic Soy Broth (TSB, Difco) containing MUG substrate (Sigma). MUG-positive isolates were presumed to be *E. coli* and were verified using the IMViC series of tests (Indole, Methyl Red, Voges-Proskauer, Citrate). Isolates exhibiting ++-- IMViC profiles were confirmed as *E. coli*

AFLP Primers and Adapters

All oligonucleotides were synthesized by Genomechanix, Inc. (Alachua, FL). The structure of the *EcoRI* adapter is:

5'-CTC GTA GAC TGC GTA CC-3'
3'-CAT CTG ACG CAT GGT TAA-5'

The structure of the *MseI* adapter is:

5'-GAC GAT GAG TCC TGA G-3'
3'-TA CTC AGG ACT CAT-5'

Primers were generated that were specific for adapter sequences and contained either one or two selective bases at their 3' end to increase their specificity and limit the number of AFLP products. Primers containing one selective nucleotide were used in the "preselective" amplification step and those containing two selective

nucleotides were used in the subsequent "selective" final amplification. The sequence of the primer sets were as follows: Selective nucleotides are in bold. (*EcoRI* 5'-GACTGCGTACC AATTC **AC**-3') (*MseI* 5'- GATGAGTCCTGAGTAA **CA**-3').

Preparation of Genomic DNA and Adapter Ligation

Genomic DNA was extracted using the DNA Easy kit according to manufacturer's instructions (Invitrogen, Inc., Carlsbad, CA). The restriction and ligation steps were then performed simultaneously according to the method of Vos et al. (1996) with minor modifications. 6.2 μL (~250 μg) of genomic DNA was then mixed with an equal volume of restriction- ligation mix, and the reaction was incubated for 3 hours at 37° C. After the reaction was completed, the mixture was diluted in 187 μL of nuclease-free water and used as a template for preselective PCR amplification.

AFLP Reactions

The AFLP reactions were performed using the primer sets described previously. PCR reactions were carried out using HotStarTaq DNA polymerase and reagents (Qiagen, Inc.). Both preselective and selective amplification steps were performed and the reaction conditions varied so as to decrease the number of AFLP fragments observed after the selective amplification. The preselective amplification was performed using the following protocol: 95° C for 15

minutes (to activate the HotStarTaq), followed by 30 cycles of 94° C for 20 s, 56° C for 30 s, and 72° C for 2 min. A final extension at 72° C was then performed for 2 min. followed by 60° C for 30 minutes. The selective amplification utilized 1 μL of the product from the preselective amplification as a template. This amplification used the same reaction conditions as in the preselective amplification; however, to increase specificity and decrease the number of AFLP products, the selective amplification used an initial annealing temperature of 66° C which was decreased one degree for nine subsequent cycles to 56° C and was followed by an additional 20 cycles using the 56° C annealing temperature. All amplification steps were performed using an Eppendorf Mastercycler Thermocycler.

Repetitive Element Polymerase Chain Reaction (rep-PCR)

Rep-PCR reactions were carried out on whole cells using the procedure described by Versalovic et al. (1994) and the BOXA1R primers described by Versalovic et al. (1998) and subsequently used by Dombek et al. (2000). The sequence of the BOXA1R primer is 5'-CTACGGCAAGGCGACGCTGACG-3'.

PCR Amplification of rRNA Intergenic Spacer Regions

The PCR ribotyping procedure was performed using the primer sets shown in Table 4. PCR reactions were carried out using HotStarTaq polymerase and reagents (Qiagen, Inc.). The amplification

was carried out using the following conditions: 95° C for 15 minutes (to activate the HotStarTaq), followed by 2 cycles of denaturation at 94° C for 1 min, 10 sec., annealing at 55° C for 2 min. 30 s, and extension at 72° C for 3 min. Cycles 3-34 included the following: denaturation at 94° C for 30 s, annealing at 55° C for 2 min 30s, and extension at 72° C for 3 min. After the final cycle was complete, the reaction was extended at 72° C for 10 min. Changing denaturing and annealing times to 20 seconds and 30 seconds, respectively, increased diversity of banding pattern and was used for subsequent reactions.

PCR Amplification of *fimA* and *fimH* Genes

Primers designed to amplify the major structural component of type 1 fimbriae in *E. coli* (*fimA*) and its flanking regions (Li et al. 1997) as well as the lectin of type 1 fimbriae (*fimH*) were developed using published sequences (Li et al. 1997; GenBank) and are shown in Table 5. PCR reactions were performed in a 20 μL reaction mixture containing 1X PCR buffer, 2.5 mM $MgCl_2$, 200 μM of each of the four deoxyribonucleotides, 0.3 μM of each primer, and 2.5 U of HotStarTaq DNA polymerase (Qiagen). Amplification was performed for 30 cycles consisting of 94°C for 1 min, 65°C for 1 min, and 72°C for 1 min., followed by 10 min at 72° C. PCR products were separated on a 1.5% agarose gel.

Table 4. Primer sets used to amplify ribosomal DNA intergenic spacer regions in *Escherichia coli*

	Primer Pair
rDNA 1	5′-AAGTCGTAACAAGGT-3′ 5′-TACTGGTTCACTATCGGTCA-3′
rDNA 2	5′-TTGTACACACCGCCCGTCA-3′ 5′-GGTACCTTAGATGTTTCAGT-3′
rDNA 3	5′-AAGTCGTAACAAGGT-3′ 5′-GGTACCTTAGATGTTTCAGT-3′

Table 5. PCR primers used to amplify the *fimA* and *fimH* genes in *E. coli*

Primer Name	Sequence	Product size (bp)
FimA fwd	5′-ACGTTTCTGTGGCTCGACGCATCT-3′	850
FimA rev	5′-ACGTCCCTGAACCTGGGTAGGTTA-3′	
FimH fwd	5′-TACCGCTATCCCTATTGGCGG-3′	480
FimH rev	5′-ACATCACGAGCAGAAACATC-3′	

PCR Amplification of *papG* Genes in *Escherichia coli*

PCR reactions to amplify *papG* genes were performed in a 20 μL reaction mixture containing 1X PCR buffer, 2.5 mM $MgCl_2$, 200 μM of each of the four deoxyribonucleotides, 0.3 μM of each primer, and 2.5 U of HotStarTaq DNA polymerase (Qiagen). An initial step of 15 minutes at 95° was performed to activate the DNA polymerase. Amplification was performed for 30 cycles consisting of denaturation (94° for 1 min.), annealing (56° for 30 s), and extension (72° for 2 min.) This was followed by a final extension at 72° for 10 min. Primer pairs specific for each of the three classes of *papG* genes were identical to those used by Johnson and Brown (1996) (Table 6).

Restriction Enzyme Analysis of rDNA, *fimA* and *fimH* Genes

Five μL of each PCR-amplified rDNA product was used for digestion with several restriction endonucleases according to the manufacturer's instructions (New England BioLabs, Inc., Beverly, Mass.). Several restriction enzymes were used: *AluI, HaeIII, EcoRI, HindIII, BamHI, Sau3AI, and MseI*. *FimA* and *FimH* PCR products (8μL) were used directly for restriction digestion with 3 U of *AluI* (New England Biolabs) without further purification at 37° C for one hour.

Table 6. Primer sets used to amplify variant *papG* genes in *Escherichia coli*

PapG class	Primer pair	product size (bp)
Class I	5′-TCGTGCTCAGGTCCGGAATTT-3′ 5′-TGGCATCCCCCAACATTATCG-3′	461
Class II	5′-GGGATGAGCGGGCCTTTGAT-3′ 5′-CGGGCCCCCAAGTAACTCG-3′	190
Class III	5′-GGCCTGCAATGGATTTACCTGG-3′ 5′-CCACCAAATGACCATGCCAGAC-3′	258

Gel Analysis

All AFLP, RFLP, and various PCR products were analyzed using either 1.5% agarose or 5% polyacrylamide gel electrophoresis. DNA was viewed using GelStar nucleic acid stain (Biowhittaker, Inc.) and UV light or silver staining, respectively. Unique bands were identified by visual analysis.

Excision of Unique Bands and DNA Recovery

Following the various fingerprinting procedures, bands unique to either human source or nonhuman source *E. coli* were identified by visual analysis. DNA from polyacrylamide gels was extracted with a Microcon purification kit (Millipore Inc.) and resuspended in nuclease-free deionized water. DNA bands from agarose gels were extracted using a gel extraction kit (Qiagen, Inc.) and prepared for cloning according to manufacturer's specifications.

Cloning of PCR products and Fimbrial Genes

Following gel extraction and DNA recovery, cloning was carried out using a TOPO TA Cloning kit (Invitrogen, Carlsbad, CA). Amplicons were ligated into the pCR 2.1 vector and transformed into competent *E. coli* TOP10F' cells. Transformed cells were plated on Luria-Bertani agar supplemented with 100 μg/mL ampicillin, isopropyl-β-D-thiogalactopyranoside (IPTG), and 5-bromo-4-chloro-3-indolyl-Beta-D-galactoside (X-gal).

Analysis of Positive Clones

The pCR 2.1 TOPO TA cloning vector is a *lacZ*- based system and analysis of positive clones was performed using blue/white screening.

Plasmid Isolation and DNA Sequencing

Plasmids were extracted using a plasmid miniprep kit (Qiagen, Inc.) according to manufacturer's instructions. Dideoxy sequencing reactions were performed using a Perkin Elmer GeneAmp PCR System 9600 (Perkin Elmer-Cetus, Norwalk, Conn.) Extension products were separated and read with a LI-COR DNA Sequencer model 4000L.

Sequence and Phylogenetic Analyses of *FimA* Gene

Ten unique *FimA* sequences from both human and animal-derived *Escherichia coli* isolates were identified in this study and were combined with data from 11 previously published *E. coli fimA* sequences (GENBANK). Isolate designations and sources are depicted in Table 7. The sequences were aligned and compared using Biowire Jellyfish software, and phylogenetic analyses were conducted using Clustalw and Treeview software. Additional *E. coli* isolates were characterized; however, some overlap was observed in general genetic sequence of the *FimA* gene. These isolates were used in confirmatory PCR tests as well as in additional ribotyping analyses.

Table 7. *Escherichia coli* isolates used in phylogenetic analyses

FimA GenBank No.	Strain I.D. designation	Host Information	Reference
None	Group I (AA1)	Human	This study
AF206652	EPEC (Group I)	Human	Peek et al. (2001)
AF206658	O157:H7 (Group I)	Human	Peek et al.
AF206659	O157:H7 (Group I)	Human	Peek et al.
U20815	O157:H7 (Group I)	Human	Unpublished
M27603	Uropathogenic (Group I)	Human	Orndorff & Falkow (1985)
X00981	K-12 (Group I)	Human	Klemm (1984)
None	PS171 (Group I)	Poultry1	This study
None	PN233 (Group I)	Poultry2	This study
AF206656	EPEC (no serogroup des.)	Human	Peek et al.
None	Group II (ET1)	Human1	This study
None	Group II (ET2)	Human2	This study
None	Group II (G1)	Human3	This study
AF206657	O55:H7 (Group II)	Human	Peek et al.
D13186	O:78 (Group II)	Avian	Sekizaki et al. (1993)
None	BC232 (Group II)	Beef	This study
None	PS175 (Group II)	Poultry	This study
None	Group III (E19)	Human	This study
Y10902	Group III	Human	Unpublished
None	Group III (DC8)	Dairy cow	This study
Z37500	O2:K1 (Group III)	Avian	Marc & Dho-Moulin (1996)

PCR Amplification of DNA Sequences Within the *FimA* Gene

Sequence analysis of the *FimA* gene revealed differences between *E. coli* isolates from human and animal sources and six separate primers (3 forward, 3 reverse) were developed that were complementary to these regions. Nested PCR reactions were performed on PCR-amplified *FimA* genes (1μL of diluted PCR product) using one of the three TMS forward primers (TMS1, TMS2, and TMS3) and three separate reverse primers, TMS*gg*,TMS*ag*, and TMS*ct* (Table 8; explanation of source of primers in results section). The reactions were carried out in 20μL volumes containing 1X PCR buffer, 200μM each dNTP, 0.3μM of each primer, and 2.5 U of HotStarTaq DNA polymerase. The cycling conditions were as follows: 95° C for 15 minutes (to activate HotStarTaq), followed by 30 cycles of 94° C for 1 min., 58° C for 1 min., and 72° C for 1 min.

Results

Collection of *E. coli* from Florida livestock and Human Sources

Fecal samples from humans and animals were identical to those collected and analyzed in the ribotyping study described previously. *Escherichia coli* were either isolated separately or isolates that had been ribotyped were analyzed for the presence of discriminating sequences. Only isolates displaying a definitive human or nonhuman

Table 8. Sequences of TMS and REV primer sets.

Primer	Sequence
TMS1 fwd	5′-ACGGCGATTGATGCGGG-3′
TMS2 fwd	5′-TACGGCAATTGATCGTACG-3′
TMS3 fwd	5′-GACTGAGATTGACAGAGCT-3′
TMS3′ fwd	5′-CTGGCTGAAACTACACCAC-3′
TMS4 fwd	5′-GTACCGCAATTGATGGCTCT -3′
REV*gg**	5′-GCAGCACCCGGGGTTGA*GG*-3′
REV*ct*	5′-GCAGCACCCGGGGTTGA*CT*-3′
REV*ag*	5′-GCAGCACCCGGGGTTGA*AG*-3′

*The last two 3′ nucleotides of the REV primers are unique. Underlined nucleotide is an intentional mismatch to increase specificity of the primer. TMS3′ fwd and TMS4 fwd are alternative primers. A 437 bp product is generated using TMS3′ fwd primer. TMS4 fwd amplifies a 212 bp product in nonhuman source isolates along with REV*gg* that are not identified by the TMS1, TMS2, or TMS3 forward primers.

ribotype profile were used; however, additional isolates were not ribotyped prior to analysis of *FimA* sequences.

AFLP reactions

AFLP products were analyzed by both agarose gel electrophoresis (Figure 4) and polyacrylamide gel electrophoresis (Figure 5). Several unique bands were identified and were purified and cloned for sequencing.

Screening of Primers Derived from AFLP Fragments

Once clones were sequenced, primers were designed that flanked the cloned regions (Table 9). Eight primer sets were developed but were not useful for differentiating sources of *E. coli*. Interesting results were obtained using one primer set (set A), however, and are shown in Figure 6. This primer set amplified several regions within the *E. coli* genome and resulted in reproducible banding patterns unique to different isolates much as has been reported using rep-PCR (Dombek et al. 2001). Although no single band was amplified using this set that was capable of discriminating *E. coli* based on source, groups of bands showed up at different frequencies from isolates from different sources. In addition, restriction digestion (*AluI*) of the PCR products generated using this primer set produced additional unique patterns that allowed for further differentiation of unknown *E. coli* isolates (Figure 7). For this reason, this primer set

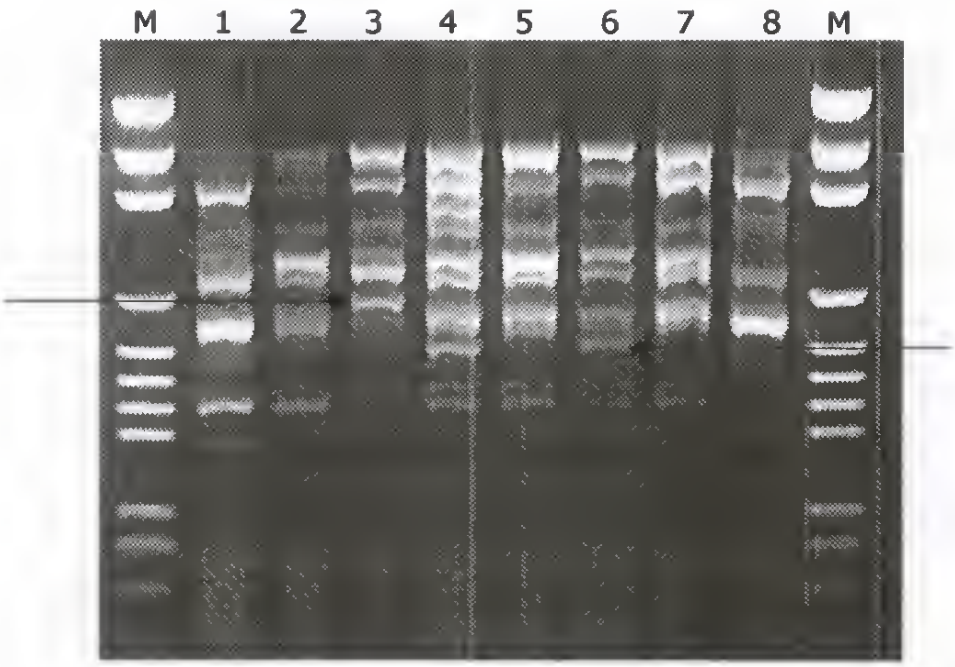

Figure 4. AFLP patterns observed in a 2% agarose gel. Lanes 1-4 are patterns from human isolates. Lanes 5-8 are patterns from nonhuman isolates. Arrows show unique bands.

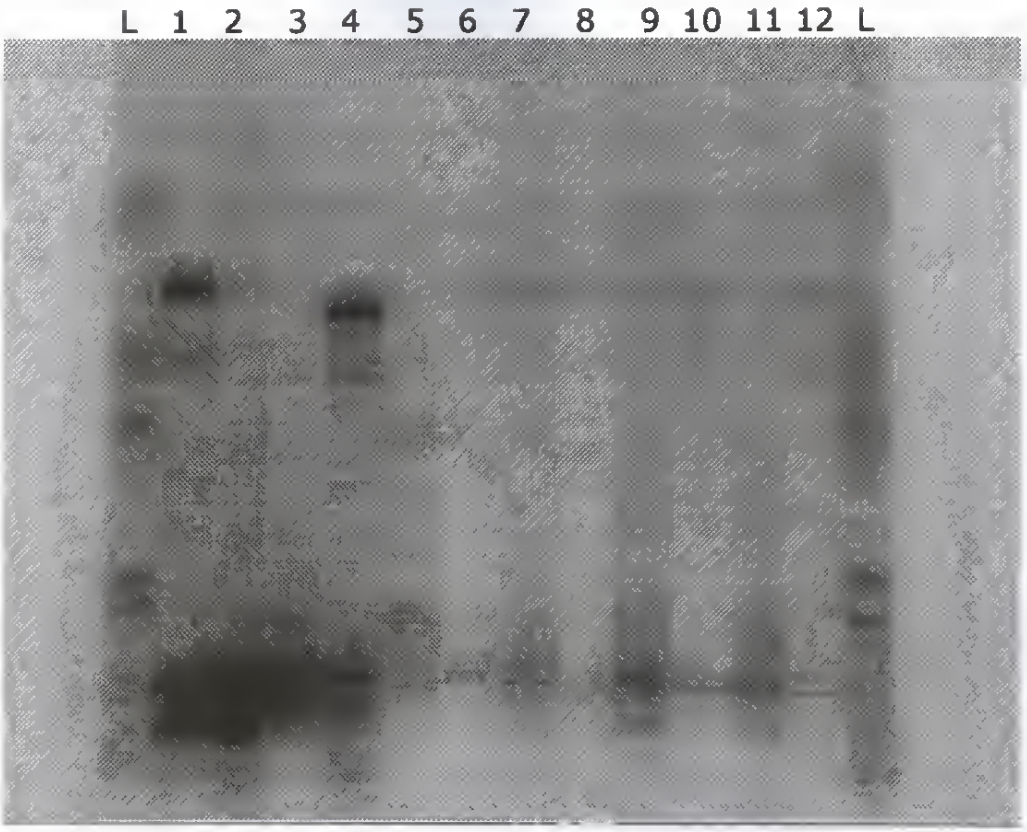

Figure 5. AFLP patterns observed on 5% polyacrylamide gel. Lanes 1,2,5,6,7,and 8 are profiles from human-derived *E. coli*. Lanes 3,4,9,10,11,and 12 are profiles from nonhuman derived *E. coli*.

Table 9. Primer sets derived from unique DNA fragments generated by Amplified Fragment Length Polymorphism

Designation	Primer Pair	Product size (bp)
Afwd	5′-GCCCCTGATTGTCCGCCTGCCGC-3′	multiple
Arev	5′-TTACACAGGTATCCCCAAAAGG-3′	
A(2)fwd	5′-CAGGTATCCCCAAAAGG-3′	362
A(2)rev	5′-TAGCCCCTGATTGTCCGC-3′	
A(3)fwd	5′-GGTTTACACAGGTATC-3′	418
A(3)rev	5′-ACGTAGCCCCTGATTG-3′	
Bfwd	5′-ATCAGATTGTCCAAAAC-3′	642
Brev	5′-ACAGGTATCCCCAAAAAG-3′	
Efwd	5′-TTACACAGGTATCCCCAAAAGG-3′	202
Erev	5′-ATCAGATTGTCCAAAAACGGGG-3′	
E(2)fwd	5′-5′-TCATCGGCAAAATGGTCC-3′	152
E(2)rev	5′-GTATTGCAGCAGGAGCGA-3′	
E′fwd	5′-GAGTCCAGGCGTTATGAATAATC-3′	254
E′rev	5′-GAATGCAGAAATGTTGATGGGC-3′	
Ffwd	5′-AGTTTGGTGCCATCG-3′	multiple
Frev	5′-GCAAAATCCCTCTCG-3′	

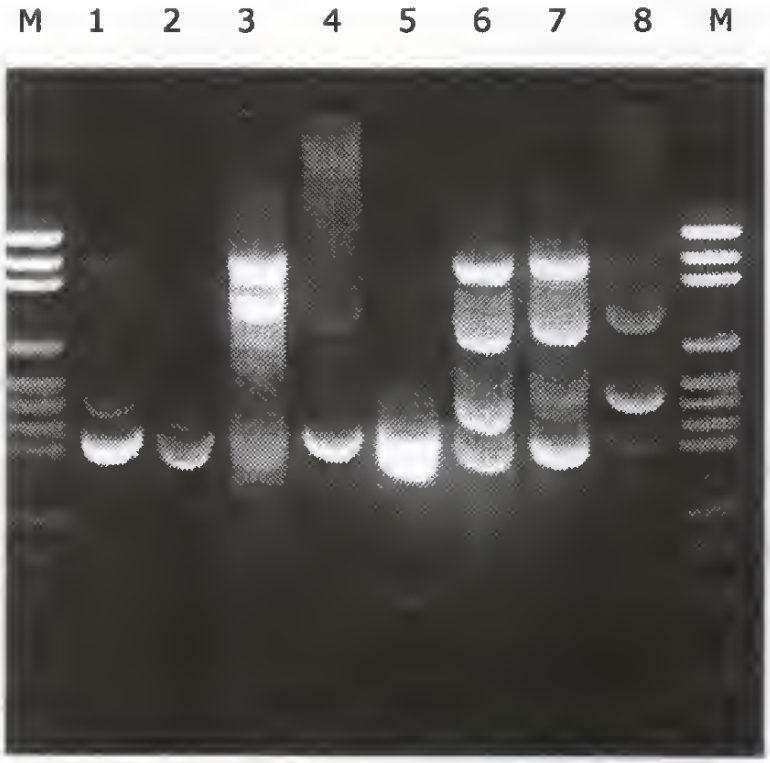

Figure 6. Rep-PCR fingerprints generated using primer set A. Lanes 1-4 are animal isolates and Lanes 5-8 are human isolates. The molecular weight marker is a pGEM *HindIII* digest (Promega, Inc.)

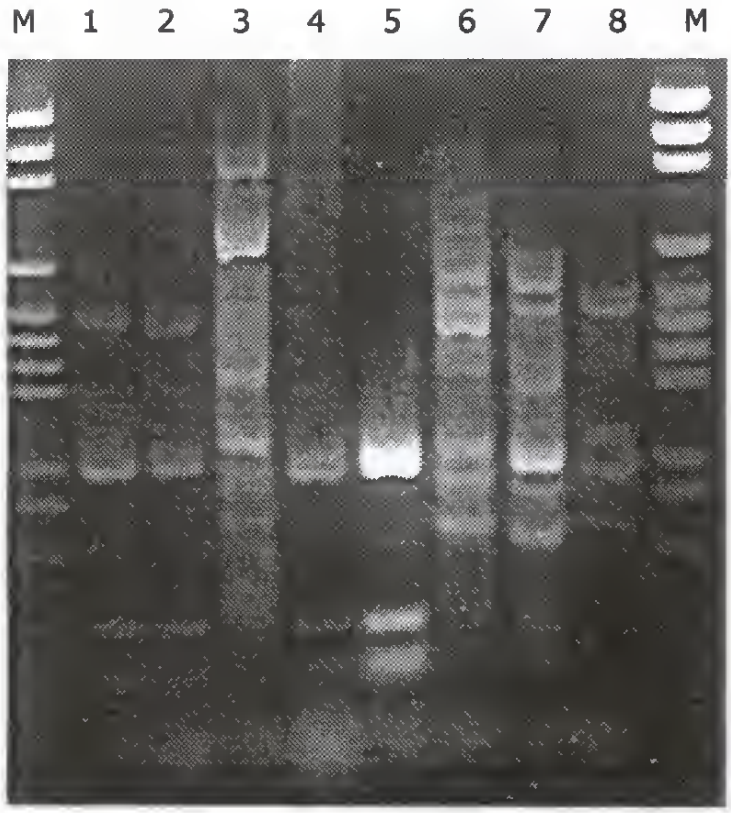

Figure 7. RFLP of rep-PCR products generated using primer set A. Lanes 1-4 are animal isolates and Lanes 5-8 are human isolates. Wells correspond to rep-PCR products in figure 6. The molecular weight marker is a pGEM *HindIII* digest (Promega, Inc.)

may prove useful as a fingerprinting primer if a reference database were to be constructed using fingerprints of isolates from known sources. This result is outside the scope of this study, as we are seeking to avoid reference-based analyses that rely on comparisons to previously characterized isolates.

Repetitive Element Polymerase Chain Reaction (rep-PCR)

Because primer A derived from a single AFLP fragment was shown to possess the capacity to amplify several regions within the *E. coli* genome resulting in multiple bands, the use of the Rep-PCR method developed by Versalovic et al. (1994) and later used by Dombek et al. (2000) was applied to *Escherichia coli* isolated from humans and livestock in the state of Florida. Several isolates were fingerprinted and showed multiple banding profiles (data not shown).

PCR Amplification of rDNA Intergenic Spacer Regions

Three primer sets were used to amplify the intergenic spacer regions between the genes that code for 16S and 23S rRNA in *E. coli*. Primer set 1 provided the most variable and discriminating patterns initially (Figure 8). Primer sets 2 and 3 provided less information to be used for discriminatory analysis. This technique, however, did not result in the identification of a single band or bands that had discriminatory power. Similar to results obtained with primer set A, a reference database would need to be constructed in order to compare

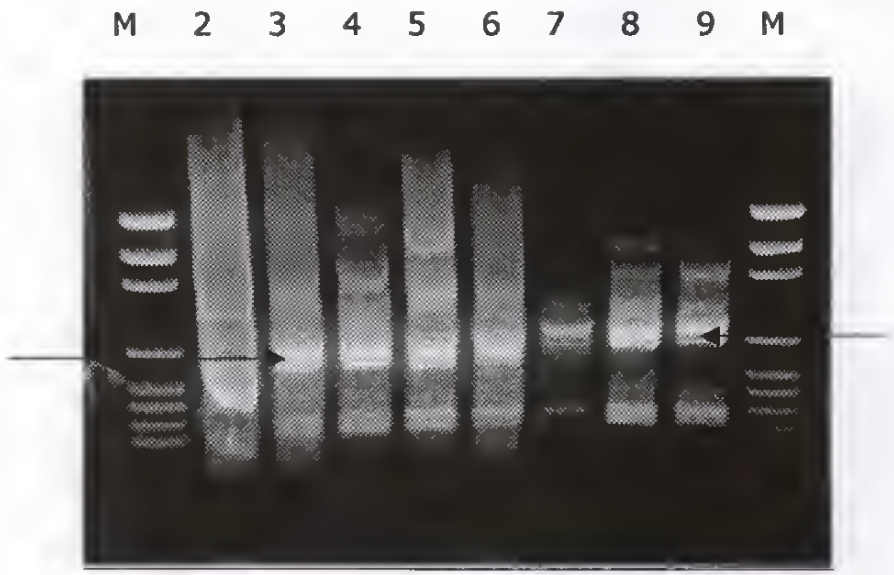

Figure 8. Amplification of rDNA intergenic spacer regions in *Escherichia coli*. Lanes 2-6 are nonhuman isolates. Lanes 7-9 are human isolates. The molecular weight marker is a pGEM *HindIII* digest (Promega, Inc.). Arrows point to doublets that are most discriminatory using this procedure. (Note that animal source isolates have lower molecular weight doublets than human source isolates).

profiles from unknown isolates to those generated using known isolates.

Cloning and Sequence Analysis of *PapG* genes in *E. coli*

PCR was used to amplify the variant *papG* genes (classes I, II, and III) of *Escherichia coli* isolated from humans (4) and animal (4) sources. Only the *papGII* allele was detected in these isolates and all eight PCR fragments were cloned and sequenced. Sequence analysis revealed no consistent or significant differences between *E. coli* isolated from humans and those isolated from animals. Additional isolates were screened for the presence of either *papGI* or *papGIII* alleles but the genes were not detected. Because of the absence of the other two *papG* alleles and the homogeneity of the *papGII* gene in these isolates, no further organisms were characterized on the basis of this gene.

PCR Amplification of *FimA* and *FimH* Genes in *E. coli*

Amplification of the *fimA* gene in *E. coli* produced a 850 bp fragment (Figure 9). Amplification of a region within the *fimH* gene produced a 480 bp fragment (Figure 10).

RFLP Profiles of PCR-Amplified rDNA and *Fim* Genes

Restriction digestion of rDNA amplicons produced several patterns, indicating the highly polymorphic nature of the ribosomal

intergenic spacer region. No single restriction pattern was unique to either human or animal isolates (Data not shown).

Restriction digestion of *FimA* genes using *AluI* are shown in Figure 11. No single RFLP pattern provided any discriminatory power, but the diversity of RFLP patterns prompted further investigation of the gene. Subsequent sequence analysis revealed that significant sequence differences are present within isolates exhibiting similar or identical RFLP patterns. Restriction digestion of the *FimH* PCR product by various enzymes failed to produce any variability in restriction patterns.

Cloning and Sequence Analysis of *FimH* Genes in *Escherichia coli*

FimH genes (*E. coli*) from human (3) and animal (2) sources were cloned and sequenced. Sequence analysis revealed no significant or consistent variability within the gene between isolates from different host sources. Minor variations were detected that resulted in changes in amino acid composition and previous data has shown that a change of a single amino acid can significantly alter lectin binding specificity (Harris et al. 2001; Schembri et al. 2001); however, these differences were not conserved within one group of *E. coli* isolates and this gene was not investigated further. Cloning and sequence analysis of *FimA* genes was discussed previously and isolates characterized in this study are listed in Table 7.

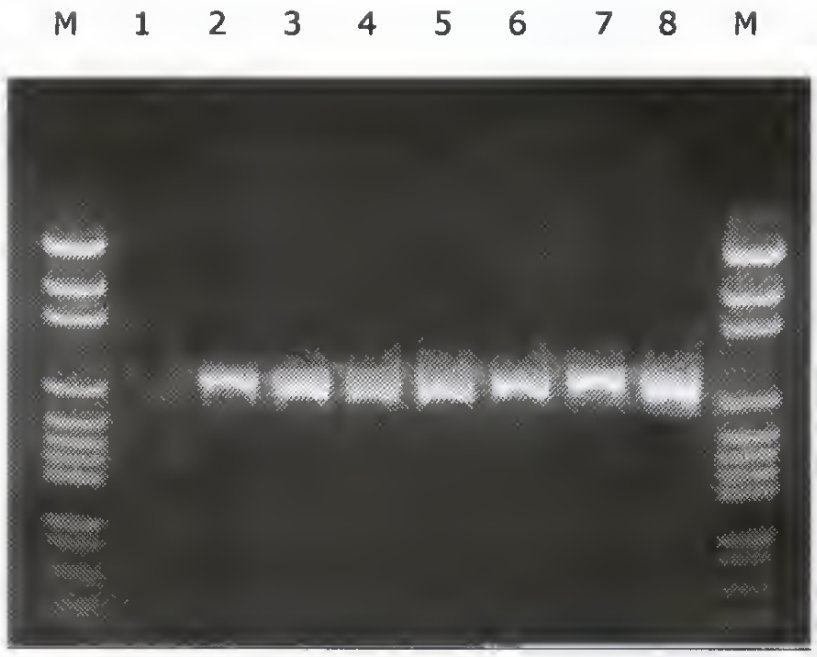

Figure 9. Amplification of the gene encoding the major structural component (*FimA*) of type-I fimbriae in *Escherichia coli*. Lane 1 is a *FimA*(-) animal isolate. Lanes 2-4 are animal isolates. Lanes 5-8 are human isolates. The molecular weight marker is a pGEM *HindIII* digest (Promega).

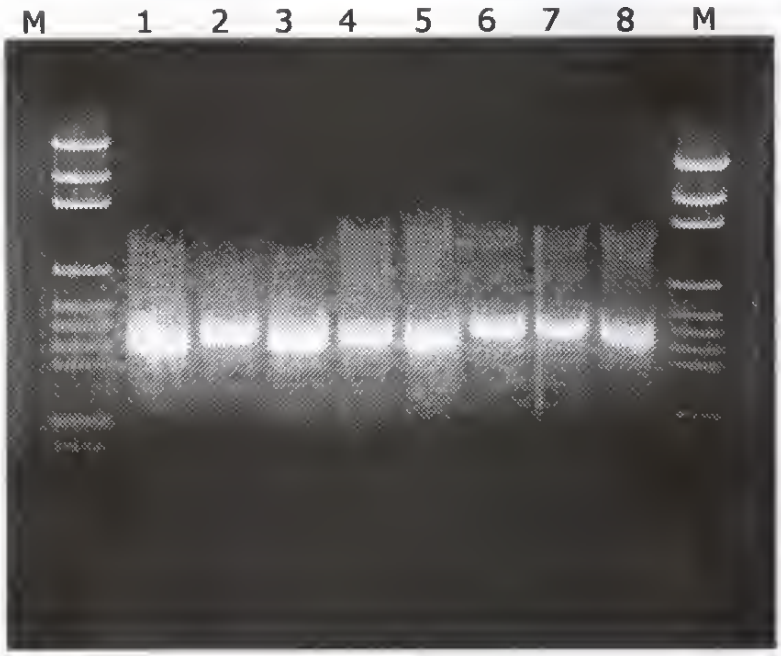

Figure 10. Amplification of the gene encoding the lectin component (*FimH*) of type-I fimbriae in *Escherichia coli* Lanes 1-4 are animal isolates. Lanes 5-8 are human isolates. The molecular weight marker is a pGEM *HindIII* digest (Promega).

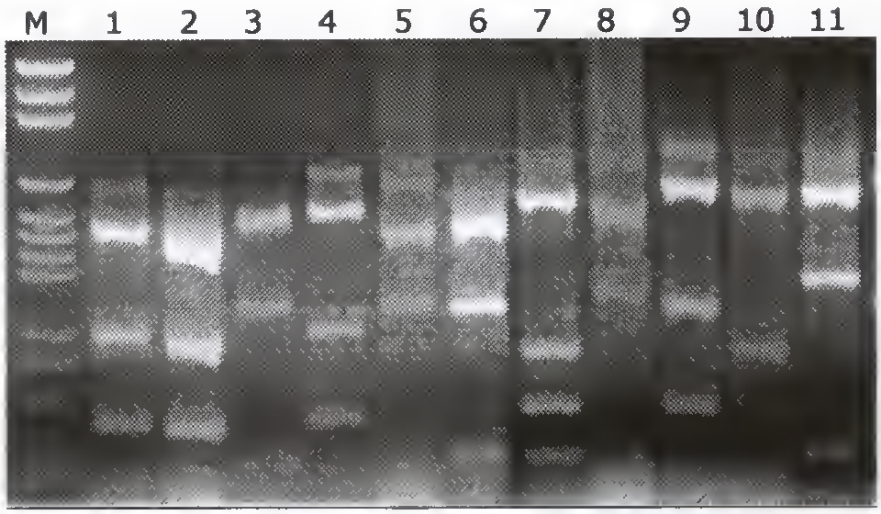

Figure 11. RFLP patterns observed after *AluI* restriction enzyme digestion of the *fimA* gene in *Escherichia coli.* Lanes 1-6 are human isolates. Lanes 7-11 are animal isolates.

DNA Alignment Analysis of *FimA* Genes

Alignment analysis of human and animal-derived *FimA* genes from Groups I, II, and III, revealed several regions of genetic diversity (Figures 12,13,14,15). Group-specific forward primers, TMS1 (bp307-312), TMS2 (bp307-312), and TMS3 (bp313-318) were complementary to the most conserved cluster of diversity. Group III isolates also contained a six base pair insertion a positions 78-83 and this area was used to construct an alternative primer set, TMS3' fwd (Table 8). Several other differences were also identified within the groups, but these differences were less conserved within organisms isolated from the same host. One region in particular, however, contained two base changes (bp495-496 in groups I and II; bp500-501 in Group II) that corresponded more consistently with host source (Figures 12-15). In group I, which contains predominantly human isolates, an adenine and guanine at positions 495 and 496 are unique to human pathogenic *E. coli* isolates (EHEC, EPEC) while two cytosines at this position are present in organisms isolated from humans, and to a lesser extent, animal isolates. In group II, the trend is reversed, and Group III isolates possess the same discriminating downstream sequences as Group I, however, their position is shifted to positions 500 and 501 due to upstream insertions. The differences at

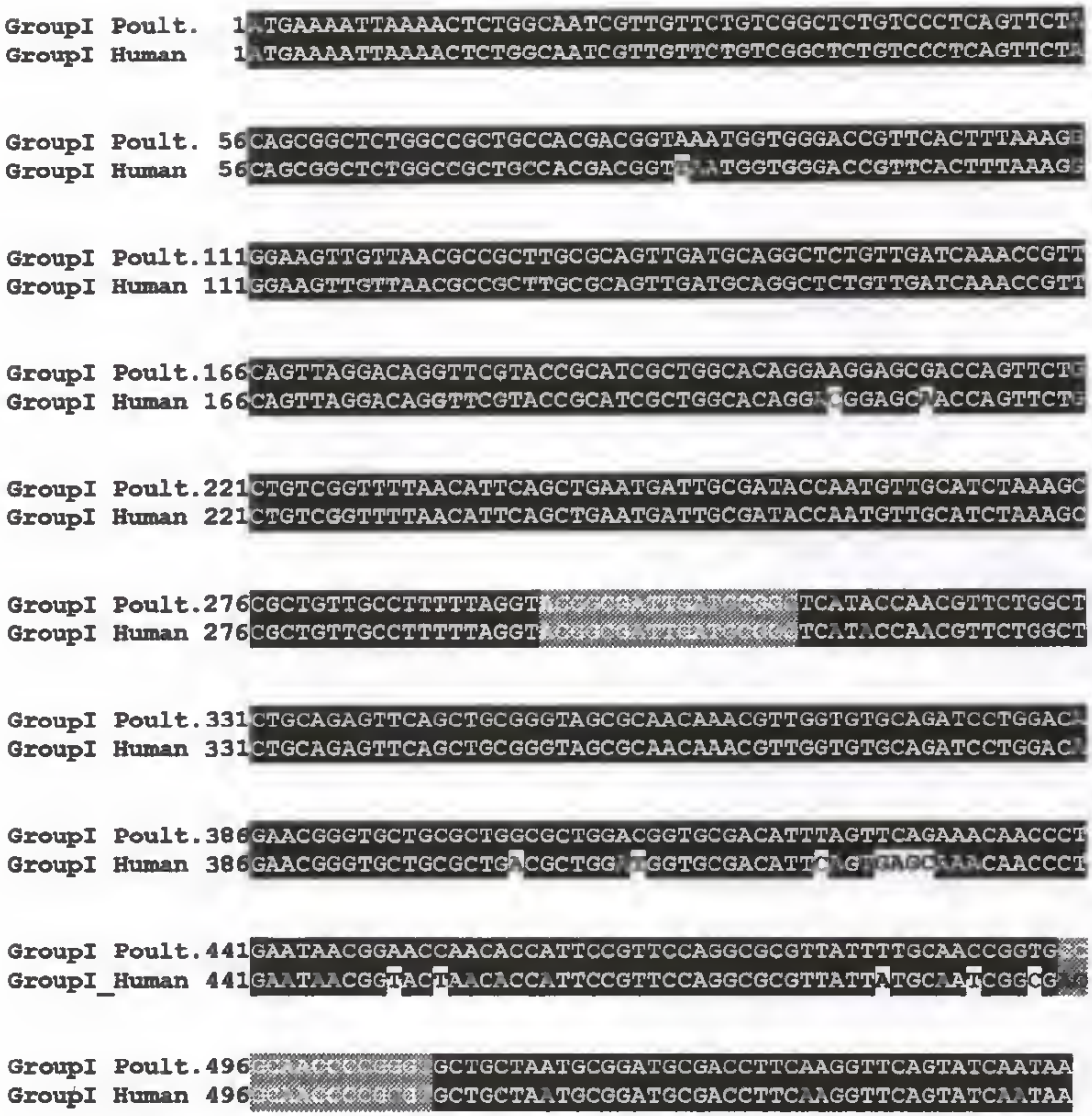

Figure 12. DNA sequence alignments of *FimA* genes from group I *E. coli* isolates. TMS and REV primers are highlighted in gray.

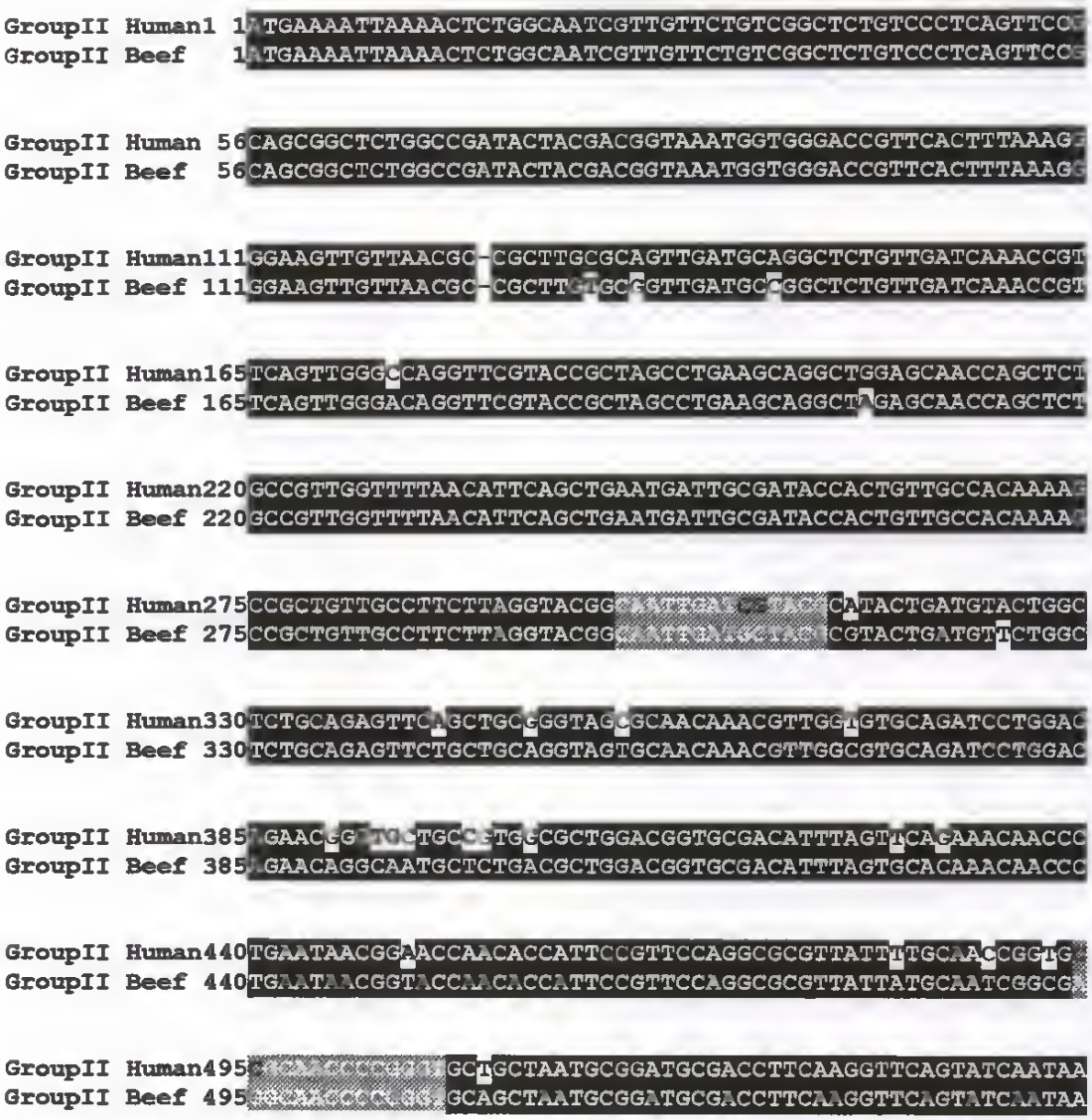

Figure 13. DNA sequence alignments of *FimA* genes from group II *E. coli* isolates. TMS and REV primers are highlighted in gray.

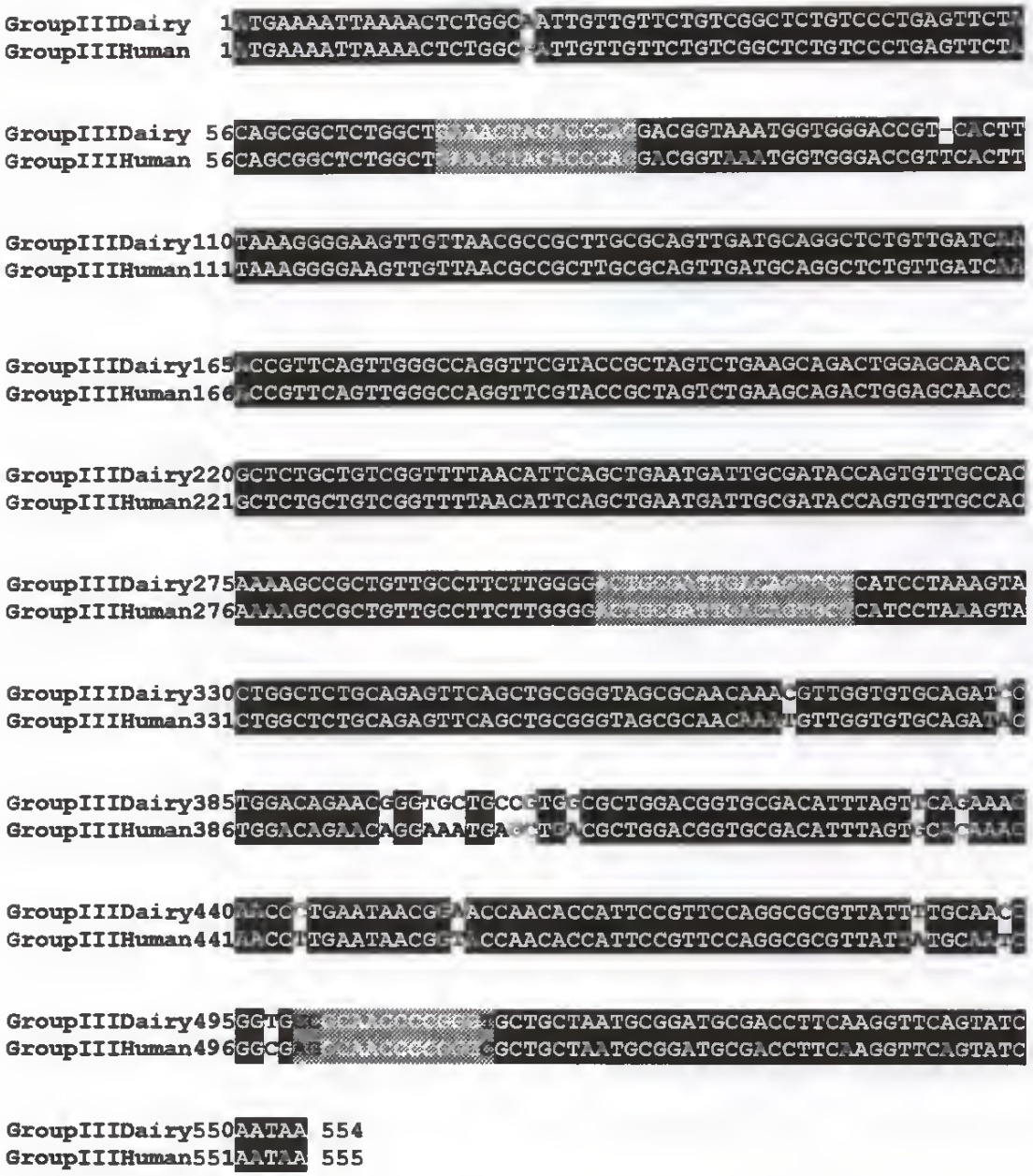

Figure 14. DNA sequence alignments of *FimA* genes from group III *E. coli* isolates. TMS and REV primers are highlighted in gray. Alternative fwd primer set is also highlighted (bp70-83).

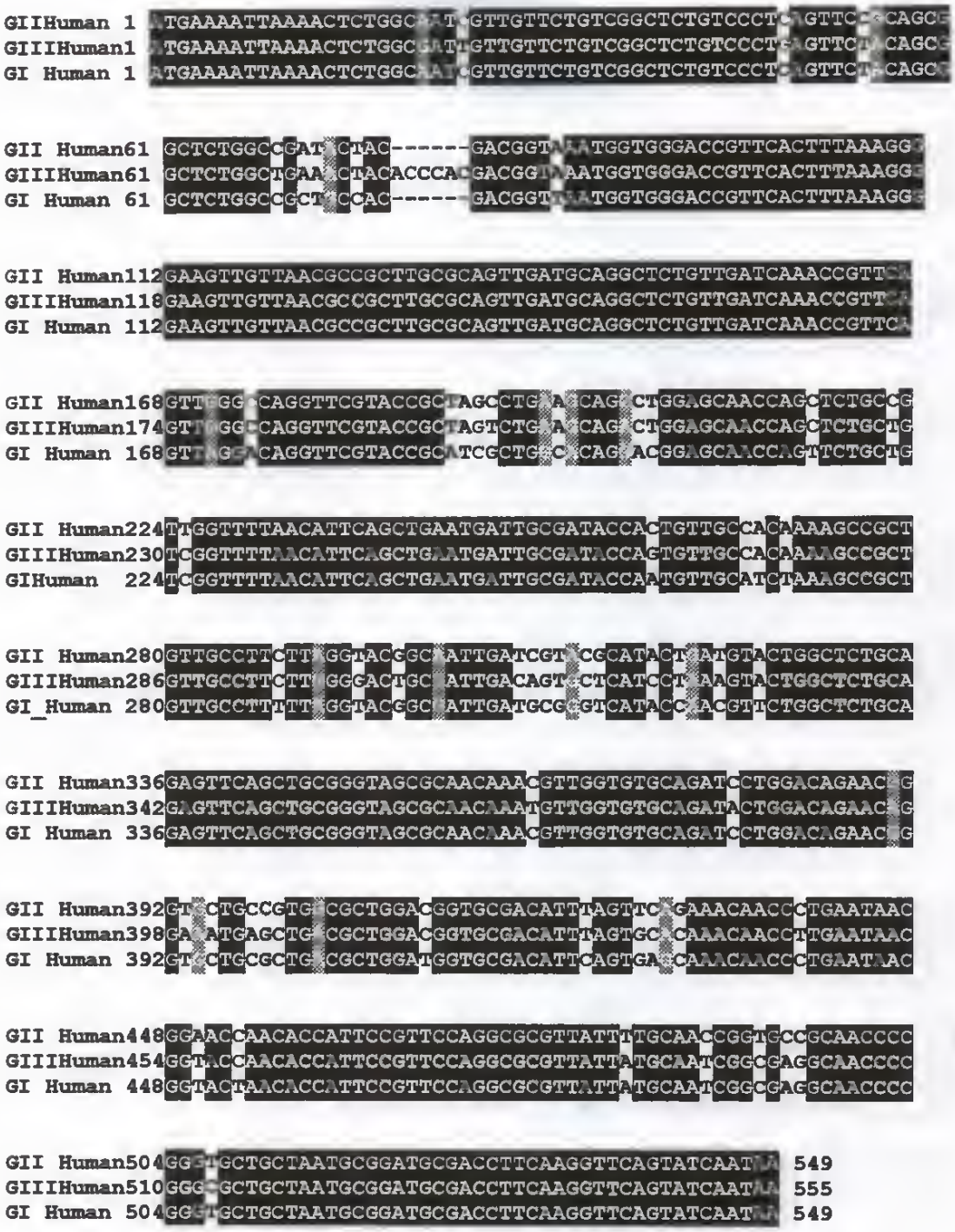

Figure 15. DNA sequence alignments of *FimA* genes from *E. coli* (groups I, II, and III).

these positions served as the template for the source-specific reverse primers REV*gg*, REV*ct*, and REV*ag* (Table 8).

Amino Acid Analysis of *FimA* Genes

Sequenced *FimA* genes were translated and analyzed using computer software (DNA Strider) in order to determine if the observed differences in DNA sequence also corresponded to phenotypic changes in the *FimA* protein. Figures 16-18 depict the amino acid sequence from groups I, II, and III, respectively. All regions within the gene that were used for primer development also contained unique amino acids. This observation supports the possibility that these changes potentially alter function and suggests a relationship between phenotype and host source. In group I, residues 103 (alanine) and 104 (glycine) were unique to this group. Likewise, group II isolates had unique amino acids (arginine, threonine) at these positions. Group III isolates had discriminating amino acids (serine, alanine) at positions 105 and 106. Furthermore, Group I isolates with a glutamic acid at position 165 are strictly human pathogens, while some overlap is observed between isolates containing an alanine at position 165. Human isolates in group II contain an alanine at position 165, while animal isolates contain a glutamic acid. Human isolates in group III contain a glutamic acid at position 167, while animal isolates in this group contain an alanine at this position.

```
1/1                                                 31/11
ATG AAA ATT AAA ACT CTG GCA ATC GTT GTT CTG TCG GCT CTG TCC CTC AGT TCT ACA gCG
Met lys ile lys thr leu ala ile val val leu ser ala leu ser leu ser ser thr ala
61/21                                               91/31
GCT CTG GCC GCT GCC ACG ACG GTT AAT GGT GGG ACC GTT CAC TTT AAA GGG GAA GTT GTT
ala leu ala ala ala thr thr val asn gly gly thr val his phe lys gly glu val val
121/41                                              151/51
AAC GCC gCT TGC GCA GTT GAT GCA GGC TCT GTT GAT CAA ACC GTT CAG TTA GGA CAG GTT
asn ala ala cys ala val asp ala gly ser val asp gln thr val gln leu gly gln val
181/61                                              211/71
CGT ACC GCA TCG CTG GCA CAG GAC GGA GCA ACC AGT TCT GCT GTC GGT TTT AAC ATT CAG
arg thr ala ser leu ala gln asp gly ala thr ser ser ala val gly phe asn ile gln
241/81                                              271/91
CTG AAT GAT TGC GAT ACC AAT GTT GCA TCT AAA GCC GCT GTT GCC TTT TTA GGT ACG GCG
leu asn asp cys asp thr asn val ala ser lys ala ala val ala phe leu gly thr ala
301/101                                             331/111
ATT GAT GCG GGT CAT ACC AAC GTT CTG GCT CTG CAG AGT TCA GCT GCG GGT AGC GCA ACA
ile asp ala gly his thr asn val leu ala leu gln ser ser ala ala gly ser ala thr
361/121                                             391/131
AAC GTT GGT GTG CAG ATC CTG GAC AGA ACG GGT GCT GCg CTG ACG CTG GAT GGT GCg ACA
asn val gly val gln ile leu asp arg thr gly ala ala leu thr leu asp gly ala thr
421/141                                             451/151
TTC AGT GAG CAA ACA ACC CTG AAT AAC GGT ACT AAC ACC ATT CCG TtC CAG GCG CGT TAT
phe ser glu gln thr thr leu asn asn gly thr asn thr ile pro phe gln ala arg tyr
481/161                                             511/171
TAT GCA ATC GGC GAG GCA ACC CCG GGT GCT GCT AAT GCG GAT GCG ACC TTC AAG GTT CAG
tyr ala ile gly glu ala thr pro gly ala ala asn ala asp ala thr phe lys val gln
541/181             ala
TAT CAA TAA
tyr gln OCH
```

Figure 16. DNA and amino acid sequence of *FimA* gene from Group I *Escherichia coli* isolates. Amino acid differences are highlighted. Residue 165 shows both amino acids present at this position in human and animal isolates.

```
1/1                                                 31/11
ATG AAA ATT AAA ACT CTG GCA ATC GTT GTT CTG TCG GTT CTG TCC CTC AGT TCC GCA gCG
Met lys ile lys thr leu ala ile val val leu ser val leu ser leu ser ser ala ala
61/21                                               91/31
GCT CTG GCC GAT ACT ACG ACG GTA AAT GGT GGG ACC GTT CAC TTT AAA GGG GAA GTT GTT
ala leu ala asp thr thr thr val asn gly gly thr val his phe lys gly glu val val
121/41                                              151/51
AAC GCC gCT TGC GCA GTT GAT GCA GGC TCT GTT GaT CAA ACC GTT CAG TTA GGC CAG GTT
asn ala ala cys ala val asp ala gly ser val asp gln thr val gln leu gly gln val
181/61                                              211/71
CGT ACC GCT AGC CTG AAG CAG GCT GGA GCA ACC AGC TCT GCC GTT GGT TTT AAC ATT CAG
arg thr ala ser leu lys gln ala gly ala thr ser ser ala val gly phe asn ile gln
241/81                                              271/91
CCG AAT GAT TGC GAT ACC ACT GTT GCC ACA AAA GCC GCT GTT GCC TTC TTA GGT ACG GCA
pro asn asp cys asp thr thr val ala thr lys ala ala val ala phe leu gly thr ala
301/101                                             331/111
ATT GAt cgT ACG CAT ACT GAT GTA CTG GCT CTG CAG AGT TCA GCT GCG GGT AGC GCA ACA
ile asp arg thr his thr asp val leu ala leu gln ser ser ala ala gly ser ala thr
361/121                                          391/131
AAC GTT GGT GTG CAG ATC CTG GAC AGA ACG GGT GCT GCG CTG GCG CTG GAC GGT GCG ACA
asn val gly val gln ile leu asp arg thr gly ala ala leu ala leu asp gly ala thr
421/141                                             451/151
TTT AGT TCA GAA ACA ACC CTG AAT AAC GGA ACC AAC ACC ATT CCG TTC CAG GCG CGT TAT
phe ser ser glu thr thr leu asn asn gly thr asn thr ile pro phe gln ala arg tyr
481/161                                             511/171
TTT GCA ACC GGT GCC GCA ACC CCG GGT GCT GCT AAT GCG GAT GCG ACC TTC AAG GTT CAG
phe ala thr gly ala ala thr pro gly ala ala asn ala asp ala thr phe lys val gln
541/181          gln
TAT CAA TAA
tyr gln OCH
```

Figure 17. DNA and amino acid sequence of *FimA* gene from Group II *Escherichia coli* isolates. Amino acid differences are highlighted. Residue 165 shows both amino acids present at this position in human and animal isolates.

```
1/1                                                 31/11
ATG AAA ATT AAA ACT CTG GCG ATT GTT GTT CTG TCG GCT CTG TCC CTG AGT TCT ACA GCG
Met lys ile lys thr leu ala ile val val leu ser ala leu ser leu ser ser thr ala
61/21                                               91/31
GCT CTG GCT GAA ACT ACA CCC ACG ACG GTA AAT GGT GGG ACC GTT CAC TTT AAA GGG GAA
ala leu ala glu thr thr pro thr thr val asn gly gly thr val his phe lys gly glu
121/41                                             151/51
GTT GTT AAC GCC GCT TGC GCA GTT GAT GCA GGC TCT GTT GAT CAA ACC GTT CAG TTG GGC
val val asn ala ala cys ala val asp ala gly ser val asp gln thr val gln leu gly
181/61                                             211/71
CAG GTT CGT ACC GCT AGT CTG AAG CAG ACT GGA GCA ACC AGC TCT GCT GTC GGT TTT AAC
gln val arg thr ala ser leu lys gln thr gly ala thr ser ser ala val gly phe asn
241/81                                             271/91
ATT CAG CTG AAT GAT TGC GAT ACC AGT GTT GCC ACA AAA GCC GCT GTT GCC TTC TTG GGG
ile gln leu asn asp cys asp thr ser val ala thr lys ala ala val ala phe leu gly
301/101                                            331/111
ACT GCG ATT GAC AGT GCT CAT CCT AAA GTA CTG GCT CTG CAG AGT TCA GCT GCG GGT AGC
thr ala ile asp ser ala his pro lys val leu ala leu gln ser ser ala ala gly ser
361/121                                            391/131
GCA ACA AAT GTT GGT GTG CAG ATA CTG GAC AGA ACA GGA AAT GAG CTG ACG CTG GAC GGT
ala thr asn val gly val gln ile leu asp arg thr gly asn glu leu thr leu asp gly
421/141                                            451/151
GCG ACA TTT AGT GCA CAA ACA ACC TTG AAT AAC GGT ACC AAC ACC ATT CCG TTC CAG GCG
ala thr phe ser ala gln thr thr leu asn asn gly thr asn thr ile pro phe gln ala
481/161                                            511/171
CGT TAT TAT GCA ATC GGC GAG GCA ACC CCG GGC GCT GCT AAT GCG GAT GCG ACC TTC AAG
arg tyr tyr ala ile gly glu ala thr pro gly ala ala asn ala asp ala thr phe lys
541/181                    ala
GTT CAG TAT CAA TAA
val gln tyr gln OCH
```

Figure 18. DNA and amino acid sequence of *FimA* gene from Groups III *Escherichia coli* isolates. Amino acid differences are highlighted. Residue 167 shows both amino acids present at this position in human and animal isolates.

Hydrophilicity Profile of Type 1 Fimbrial Subunit (*FimA*) Protein

Because genetic differences within the *FimA* gene were clustered together, the hydrophilicity profile for the *FimA* protein was calculated using the method of Hopp et al. (1981) in order to identify regions that may be exposed to the environment and would therefore be more apt to be affected by diversifying selection. The hydrophilicity profile is shown in Figure 19. As would be expected, the N-terminal region is highly hydrophobic, which is characteristic of a region harboring the signal peptide for a membrane-bound structure. Several hydrophilic maxima are also evident, with the most prominent one encompassing residues 101-107. This peak corresponds to the polymorphic region within *FimA* that is the basis of the TMS forward primer sets. Other minor peaks correspond to additional areas exhibiting additional genetic variation within the *FimA* gene.

Phylogenetic Analysis of *FimA* Genes

Phylogenetic analysis of *FimA* genes cloned in this study are shown in Figures 20 and 21. As can be seen in the Figures, the three groups cluster well in the unrooted tree analysis. The more complex rooted analysis, using the isolates cloned in this study as well as other published *FimA* sequences, also shows the groups clustering together, and a separation of human and animal isolates.

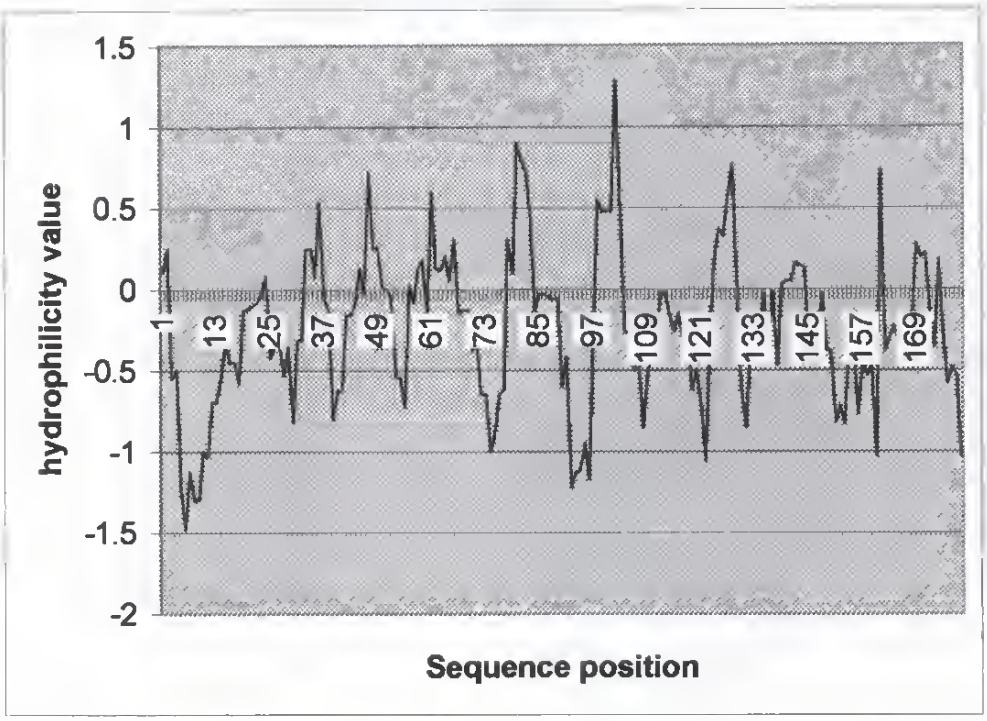

Figure 19. Hydrophilicity profile of type 1 fimbrial subunit protein (*FimA*) in *Escherichia coli*.

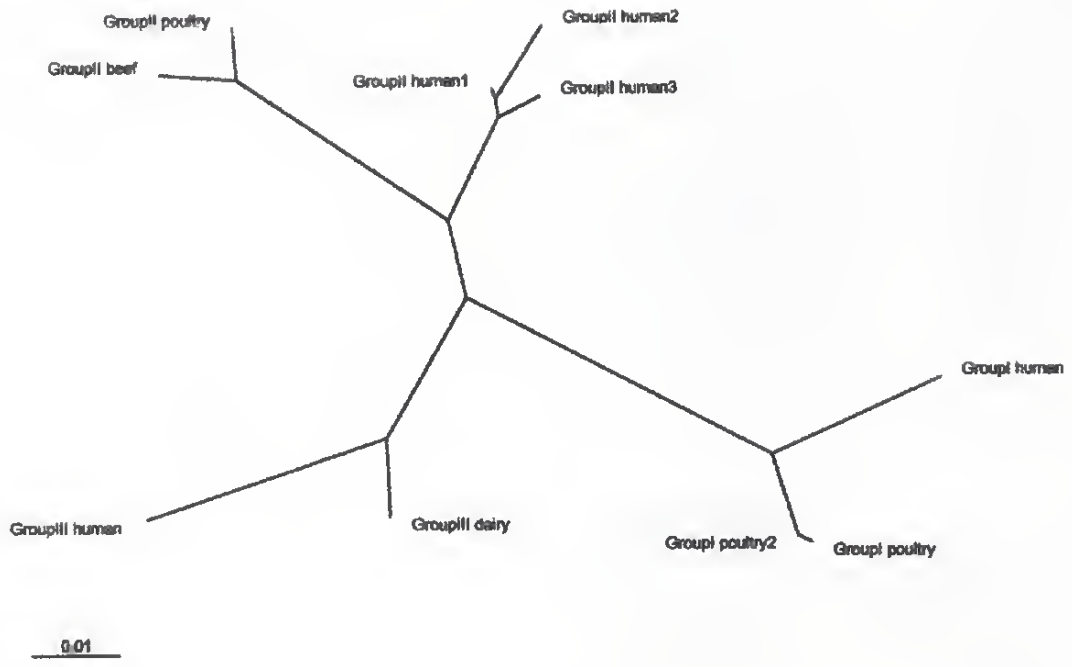

Figure 20. Unrooted phylogenetic tree showing genetic relationship between *FimA* genes from human and animal *Escherichia coli* isolates sequenced in this study. Scale bar = 1% divergence.

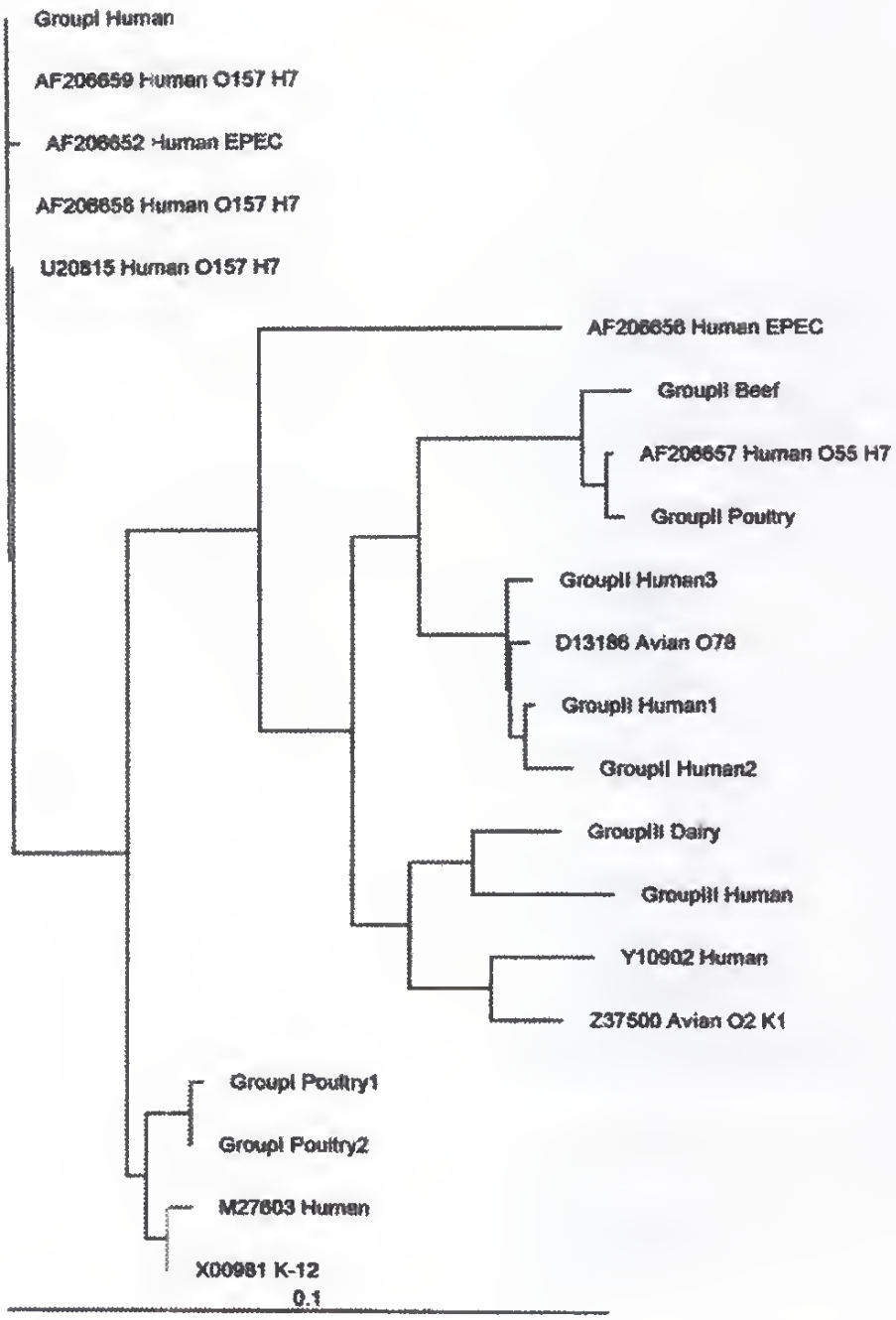

Figure 21. Phylogenetic tree showing genetic relationship between *FimA* genes from human and animal *Escherichia coli* isolates sequenced in this study and in published literature. Scale bar = 10% divergence.

Identification of *E. coli* Isolates by TMS and REV Primer Sets

Figure 22 shows a schematic of how the TMS and REV primers are used to identify the source of unknown *E. coli* isolates using a PCR-amplified *FimA* gene as template. A positive result is a 212 bp product The results of screening several *E. coli* isolates from various sources are shown in Table 10. For identification of a single isolate, nine separate PCR reactions must be performed. Each of the three forward TMS primers, (TMS1, TMS2, and TMS3), must be used separately with each of the reverse primers, REV*gg*, REV*ct*, and REV*ag*. A negative PCR result is considered indicative of a nonhuman isolate. Some isolates that cannot be identified using the three TMS forward primers can be identified using an alternative primer combination (TMS4, REV*gg*; Table 8). A positive PCR product is considered as a nonhuman *E. coli* isolate. This fimbrial sequence was not observed in this study, however. In addition, no isolates in this study were amplified using the REV*ag* reverse primer. The TMS4 forward primer and REVag reverse primer were developed from fimbrial sequences obtained from GENBANK and published literature (Peek et al. 2001). The overall correct classification rate of isolates in the Florida collection is 70%. Beef, Dairy, and Swine isolates were correctly identified as nonhuman at a rate of 68%, 79%, and 96%, respectively. Poultry isolates were only positively identified at 55% of the time, however. The reasons

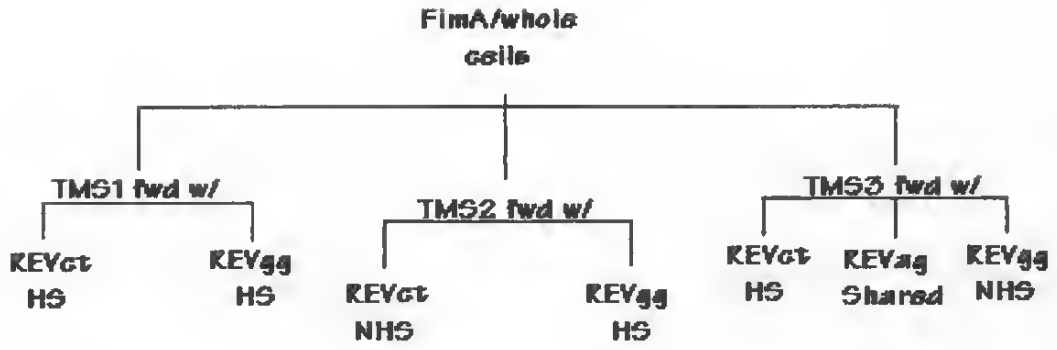

Figure 22. Schematic representation of sequence of primer usage to identify Human and Animal-derived *FimA* genes in *Escherichia coli*. *HS = human source isolate; NHS = nonhuman source isolate

Table 10. Human/nonhuman classification of *Escherichia coli* using TMS and REV primer sets

Source (No. of isolates)	No. (%) of isolates classified as: Human	Nonhuman
Beef * (25)	8 (32)	17 (68)
Dairy (56)	12 (21)	44 (79)
Swine (45)	2 (4)	43 (96)
Poultry (38)	17 (45)	21 (55)
Total Human (104)	65 (63)	39 (37)
Total Nonhuman (164)	39 (23)	125 (77)

* TMS primers do not differentiate between species. Absence of a PCR product is classified as nonhuman.

for this observation are not known; however, diet of the host animal as well as variation in intra- and extra intestinal environments may play a role. Isolates from domestic dogs were also often misclassified as being from human origin (data not shown). This observation was not surprising, however, as it would be reasonable that sharing of flora would occur within a household. For this reason, *E. coli* isolated from dogs were not further investigated as part of this study. Domestic dogs have been shown to contain *E. coli* that possess additional fimbrial types thought previously to occur predominantly in humans, however (Feria et al. 2001; Johnson et al. 2001; Yuri et al. 1998). Overall correct classification of published *FimA* sequences and those found in GENBANK is 68% (Peek et al. 2001).

Characterization of Plasmids Containing Portion of *FimA* gene

In order to investigate the possibility that Type-1 fimbrial genes could be exchanged via horizontal transfer, several plasmids were extracted from environmental *E. coli* isolates and were analyzed for the presence of fimbrial genes or gene fragments. Two plasmids (approx. 2700 bp) extracted from two environmental *Escherichia coli* isolates (one human, one beef) were designated as plasmids pFA01 and pFA02 and were used as templates for PCR amplification using various primer sets complementary to regions within the *FimA* gene. Using the TMS1, REV*gg* primer set, a 213 bp segment of the *FimA*

gene was amplified on each plasmid (Figure 23). The PCR amplicons were sequenced and were found to be completely homologous to the Group I chromosomal *FimA* allele of *Escherichia coli* (Figure 24). Subsequent PCR reactions using *FimA*-complementary primers failed to amplify additional regions within the plasmids, suggesting that only a portion of the gene is present. In addition, while pFA01 was present in both the human and animal isolate, only the human isolate possessed a Group I chromosomal *FimA* allele, while the animal isolate possessed a Group II chromosomal *FimA* allele.

Discussion

This study employed several molecular methods to identify unique DNA sequences in *Escherichia coli* that differentiate organisms isolated from human and animal sources. Several direct fingerprinting methods were used, including Amplified Fragment Length Polymorphism (AFLP) analysis, PCR ribotyping, and Repetitive Element Polymerase Chain Reaction (rep-PCR). These methods analyze the whole genome of *E. coli* and amplify multiple regions within the chromosome, thus generating a DNA fingerprint. While the use of AFLP and PCR ribotyping for this purpose is novel, rep-PCR technology has been used successfully to differentiate *E. coli* on the basis of host-source (Dombek et al. 2000). All three methods are likely useful candidates for purposes of bacterial source tracking, however, all

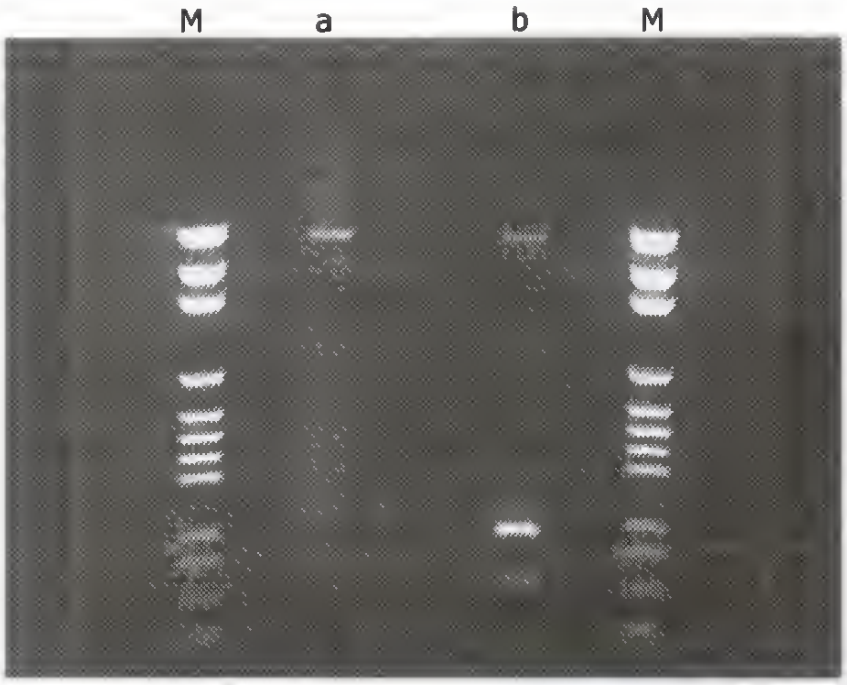

Figure 23. Gel electrophoresis of plasmid pFA01 (a) and PCR product (pCR*Fim*A) (b) using TMS1 and REV*gg* primer set. Identical results were obtained with plasmid pFA02.

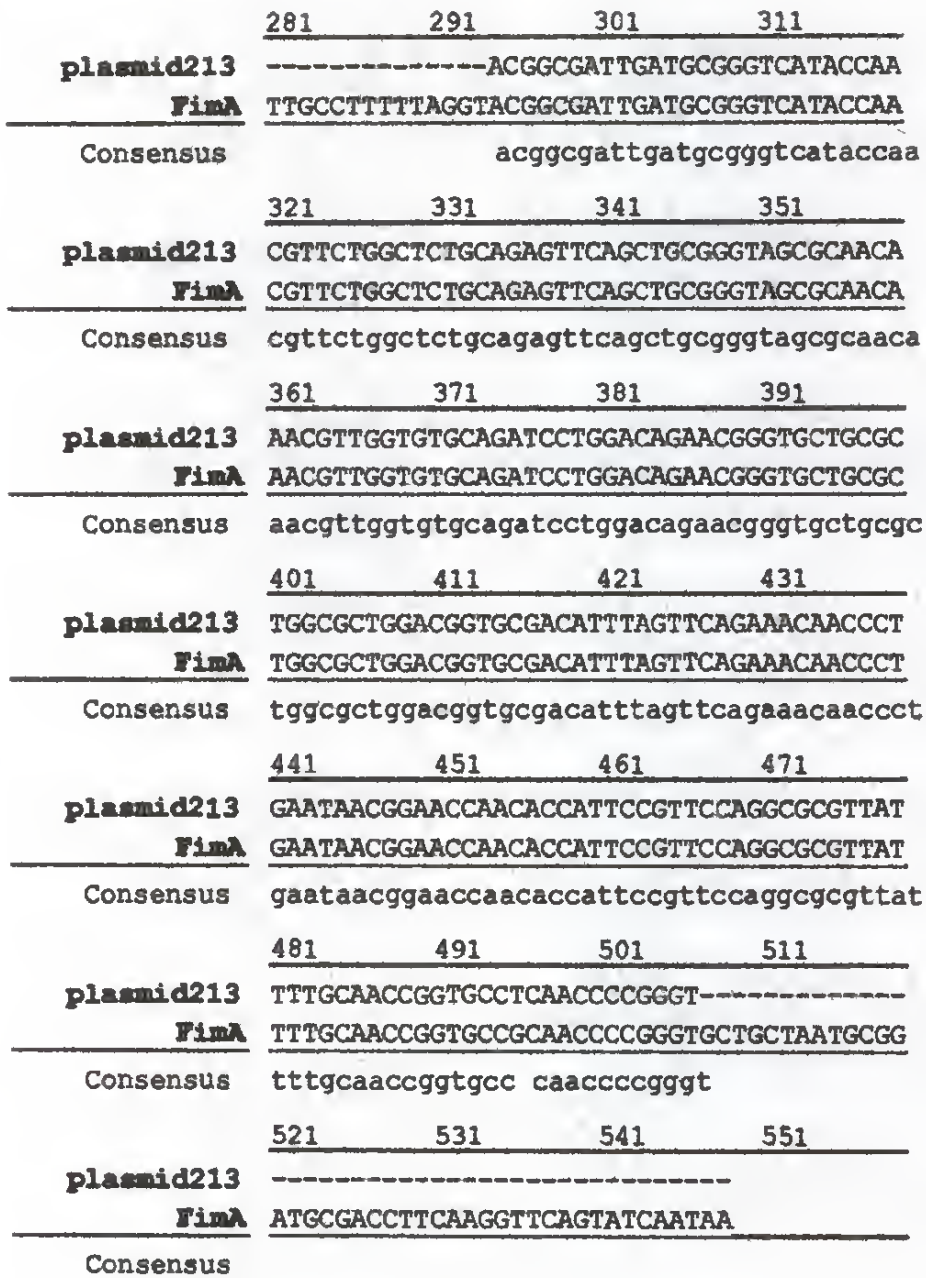

Figure 24. Sequence alignment of pCR*Fim*A and Group I *FimA* gene.

require the formation of a reference database to which fingerprints of unknown isolates can be compared. The purpose of this study was to identify specific regions within the *E. coli* genome that are unique to isolates of human or animal origin rather than to generate unique DNA fingerprints to which known isolates can be compared. The fingerprinting techniques employed in this study did not identify any such sequences. For this reason, additional methods were used, including cloning and sequence analysis of the *FimA* gene in *E. coli*, which codes for type 1 fimbriae, *FimH*, which codes for the adhesin component of type 1 fimbriae, and the three alleles of *papG*, which code for the variant adhesin components of P fimbriae. The *FimA* gene provided the most information in the form of subtle genetic differences between bacteria isolated from animals and those isolated from humans; however, further investigation is warranted on additional fimbrial genes as to genetic structure, prevalence, and epidemiology, as well as in their potential use for microbial source tracking. It was these differences that served as templates for the development of PCR primers used to amplify regions of DNA specific to *Escherichia coli* isolated from either human or animal sources. Analysis of the source of diversification of the *FimA* gene suggested that the genetic differences were the result of both diversifying selection and horizontal transfer of genes between bacteria. Overall, the results of the study

produced a molecular tool capable of aiding in differentiating *Escherichia coli* from human and animal sources; however, this gene was detected in all animal species that were sampled and evidence suggests that it may be possible that the gene can be horizontally transferred between organisms from different origins. Therefore, this method would be most useful when coupled with one or more additional bacterial source tracking methods as well as for situations involving relatively recent fecal contamination. Overall, correct classification of isolates was over 70%, making this a reliable method for predicting sources of *Escherichia coli* in the environment.

CHAPTER 4
SUMMARY AND CONCLUSIONS

Overall, this study provided insight into the possibility that unique DNA sequences are present within the genome of *Escherichia coli* that are capable of discriminating them on the basis of being derived from either humans or animals.

The results of the first half of this study indicate that the ribotyping procedure may not be an effective tool for identifying specific (species level) sources of *Escherichia coli* from different animals in different geographic locations. Although the reason *E. coli* isolates from different hosts exhibit unique ribotype profiles is not known, it has been suggested, and it is likely, that the intestinal microenvironment within the host species may play a role. This would help to explain the consistencies observed when isolates are collected from individuals within a confined region in which fluctuations in pH, temperature, and diet are likely to be minimal. When isolates are collected from a broad geographic region, such as was done in this study, external factors become more of an influence, a greater overlap is observed in RT profiles, and a lower rate of correct classification results. The results of this study do indicate, however, that ribotyping

is still an effective tool in discriminating *E. coli* from human and nonhuman sources even over a broad geographic range.

Several additional molecular approaches were also utilized to determine if unique sequences exist within the *Escherichia coli* genome that are capable of identifying fecal pollution as being from either human or animal sources.

The purpose of the second half of this study was not to establish a database of genetic fingerprints from known sources using various techniques that could, in turn, could be cross-referenced in order to classify an isolate of unknown origin. This approach has already been established as moderately useful using rep-PCR and ribotyping. While these methods have proven useful in published studies (Carson et al. 2001; Dombek et al. 2001; Parveen et al. 1999), the ribotyping results discussed previously indicate that these patterns may only be useful within a confined geographic area or watershed. Instead, the purpose of this study was to identify specific sequences within the *E. coli* genome that were unique to isolates from either human or animal origin. Several DNA fingerprinting methods were used to identify genetic differences and, rather than establishing a database containing the entire fingerprint, unique bands were purified, cloned, and subjected to sequence analysis to determine if they contained unique DNA sequences. By using fingerprinting techniques that utilize

restriction enzyme digestion, the presence or absence of a single band can be the result of a single base change that either creates or destroys the sequence recognized by a particular enzyme. While single nucleotide polymorphisms (SNPs) could potentially be used to discriminate between similar strains of bacteria (Butler and Reeder 2001; Gu et al. 1998; Watanabe et al. 2001), this approach was not investigated in this study.

The first technique utilized as a fingerprinting tool to discriminate sources of *E. coli* was amplified fragment length polymorphism (AFLP). This method proved useful in identifying bands that were unique to either human or animal-derived *Escherichia coli*. After the bands were extracted and cloned, primers were developed that were complementary to the DNA regions flanking the ends of the fragment. Subsequent PCR analysis of known-source isolates of *E. coli* showed differential amplification of some but not all isolates. The primers either failed to amplify a region within the genome of a particular isolate, or the band sizes were of varying lengths depending on the isolate. After the development of several primer sets, it was determined that, while this particular approach may have some merit, the result would likely be that several multiplex PCR reactions would be required to identify an isolate from a particular source. While this

approach still eliminates the need for establishing a reference database, the method would be time-consuming and subject to error.

PCR-ribotyping has been used previously to discriminate between similar or nearly identical organisms (Cartwright et al. 1995; Kostman et al. 1992; Matar et al. 1993). Primer sets that are complementary to conserved DNA sequences that code for ribosomal RNA (rDNA) are used to amplify a region of DNA that connects the coding regions for 16S and 23S rRNA. While the rRNA products are highly conserved throughout prokaryotes, these intergenic spacer regions are highly polymorphic. The result is the amplification of up to seven bands (in *Escherichia coli*) of varying lengths that are unique to a particular strain. This technique is very reproducible and capable of generating several unique rDNA fingerprints. Initial experiments using this method were promising as the fingerprints produced were unique to either human or animal isolates. As the number of isolates analyzed increased, more profiles were observed, and it became apparent that this method would likely require the establishment of a PCR-ribotype database in order to correctly classify an isolate of unknown origin. This method should not be ruled out as a possible approach to be used in conjunction with a reference database and statistical analysis as the procedure is inexpensive, reproducible, and requires only a single PCR reaction to produce results as compared to

the conventional ribotyping procedure which is both costly and time-consuming.

Characterization of the *FimA* gene by PCR amplification, RFLP, cloning, and sequencing resulted in the development of PCR primers capable of discriminating *E. coli* on the basis of being derived from either human or animal sources. Although the reason as to why these fimbriae are genotypically different according to host are not known, it has been speculated that genotypic (and subsequently phenotypic) differences in fimbrial adhesins are a result of adaptations to differences in the microenvironments present in different host intestines including temperature, pH, diet, and the presence or absence of fimbrial receptors. In addition, it has been suggested that because fimbriae often act as antigens recognized by the host immune system, the ability to alter the phenotype of such structures would not only allow for selection to maintain or enhance function, but would also allow for evolution of novel epitopes to avoid recognition by host defenses (Tewari et al. 1993). Analysis of the *FimA* gene product amino acid composition revealed several hydrophilic peaks that likely correspond to regions that are accessible to the environment and may also serve as antigenic sites. Interestingly, these regions were also those that showed high genetic variability, which strengthens the hypothesis that natural diversifying selection plays a role in this

observed variability. *E. coli* fimbriae (and the gene that codes for them) have long been known to be polymorphic, and several investigators have utilized these differences to classify specific groups of *Escherichia coli*. Li et al. (1997) reported the ability to use PCR primers specific for regions of the *FimA* gene to identify *Escherichia coli* O157:H7. Similarly, Marc et al. (1996) used PCR to identify differences in *FimA* to discriminate avian pathogenic and nonpathogenic *Escherichia coli*. Horizontal gene transfer has also been suggested as a mechanism of diversity and is evidenced by individual *FimA* alleles possessing gene fragments identical to two separate *Escherichia coli* isolates (Peek et al. 2001; this study). The current study also revealed that environmental isolates showed a more diverse collection of genotypes than individual fecal samples. This would support the possibility that environmental isolates may have the ability to transfer the *FimA* gene or fragments of the gene by horizontal transfer mechanisms such as conjugation. This hypothesis was greatly supported by this study, which identified a 213 bp segment of the *FimA* gene on plasmids isolated from two environmental (one human-derived, one animal-derived) *E. coli* isolates. The gene segment was amplified using the primer sets that were designed in this study (TMS1 fwd, REV*gg*). While the individual plasmids extracted from the human and animal environmental isolates contained DNA homologous to the

Group I *FimA* allele, only the human isolate possessed a Group I chromosomal *FimA* gene. The animal isolate possessed a Group II allele. This suggests that the DNA contained within the plasmid did not originate from the chromosome of the host organism and was likely acquired via some other mechanism such as horizontal transfer between unrelated *E. coli* strains.

The results of this study emphasize the potential use of the *FimA* gene in classifying strains of *E. coli*. After analysis of several published *FimA* sequences as well as those characterized in this study, it was apparent that, although different fimbrial alleles were present in all of the animal species tested, alleles were present at different frequencies within clusters of organisms that were isolated from different animal and human hosts. Not surprisingly, these genetic differences also result in changes in primary amino acid sequence and, perhaps, alter function of the fimbriae themselves.

These genetic differences were subsequently used as templates for the development of specific PCR primers that are capable of amplifying regions within the *E. coli* genome that are more prevalent in organisms derived either from human or animal sources. The result is a rapid, PCR test that can aid in differentiating *Escherichia coli* as being either human or animal derived. Correct classification rates were above 70%, which makes this method a viable addition to any

program that currently implements bacterial source tracking methodology. When used in conjunction with additional bacterial source tracking methods, this tool can be useful for the purpose of more accurately identifying sources of fecal pollution. In addition, this study suggests the possibility that different individual genes or collections of genes, especially those that code for fimbrial and nonfimbrial adhesins and outer membrane proteins, may be present in *E. coli* from different animal hosts and should be investigated further as a means of tracking the source of bacteria in the environment. The result of successful microbial source tracking is an ability to more accurately determine possible risks to human health as well as to increase the ability to remedy existing pollution problems in the environment.

REFERENCES

Atmar RL, Estes MK. Diagnosis of noncultivatable viruses, the human caliciviruses. Clin Microbiol Rev 2001;14(1):15-37.

Aznar R, Ludwig W, Schleifer KH. Ribotyping and randomly amplified polymorphic DNA analysis of *V. vulnificus* biotypes. Syst Appl Microbiol 1993;16:303-9.

Barenkamp SJ, Munson RS Jr., Granoff DM. Subtyping isolates of *Haemophilus influenza*e type b by outer-membrane protein profiles. J Infect Dis 1981;143:668-76.

Bernard AE, Field KG. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl Environ Microbiol 2000;66(4):1587-94.

Bertin YJ, Girardeau P, Darfeuille-Michaud A, Martin C. Epidemiological study of pap genes among diarrheagenic or depticemic *Escherichia coli* strains producing CS31A and F17 adhesins and characterization of Pap(31A) fimbriae. J Clin Microbiol 2000;38(4): 1502-9.

Blostein, J. Shigellosis from swimming in a park pond in Michigan. Public Health Rep 1991;106(3):317-22.

Boyd EF, Hartl DL. Diversifying selection governs sequence polymorphism in the major adhesin proteins FimA, PapA, and SfaA of *Escherichia coli*. J Mol Evol 1998;47:258-67.

Brinton CC. The structure, function, synthesis, and genetic control of bacterial pili and a molecular model for DNA and RNA transport in gram negative bacteria. Trans NY Acad Sci 1965;27:1003-54.

Burkhardt MR. Determination of submicrogram-per liter concentrations of caffeine in surface water and groundwater samples by solid-phase extraction and liquid chromatography. Journal of AOAC International 1999;82:161-6.

Butler JM, Reeder DJ. Detection of DNA polymorphisms using PCR-RFLP and capillary electrophoresis. Methods Mol Biol 2001;163:49-56.

Carson A, Shear BL, Ellersieck MR, Asfaw A. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. Appl Environ Microbiol 2001;67(4); 1503-7.

Cartwright CP, Stock F, Beekmann S, Williams EC, Gill VJ. PCR amplification of rRNA intergenic spacer regions as a method for epidemiologic typing of *Clostridium difficile*. J of Clin Microbiol 1995;33(1):184-7.

Cho HB, Lee SH, Cho JC, Kim SJ. Detection of adenoviruses and enteroviruses in tap water and river water by reverse transcription multiplex PCR. Can J Microbiol 2000;46(5):417-24.

Clesceri, LS, Greenberg AE, Trussell RR. editors. Standard Methods for the Examination of Water and Wastewater. APHA, AWWA – WPCF; 1998.

Cook GC. Liver involvement in systemic infection. Eur J Gastroenterol Hepatol 1997;9(12):1239-47

Cooke MD. Antibiotic reseistance among coliform and fecal coliform bacteria isolated from sewage, seawater, and marine shellfish. Antimicrob Agents Chemother 1976;9:879-84.

Curry A, Smith HV. Emerging pathogens: Isospora, Cyclospora, and microsporidia. Parasitology 1998;117:143-59.

Da Silva AJ, Slemenda SB, Visvesvara GS, Schwartz DA, Wilcox CM, Wallace S, et al. Detection of Septata intestinalis (Microsporidia) Cali et al.1993 using polymerase chain reaction primers targeting the small subunit ribosomal RNA coding region. Mol Diagn 997;2(1):47-52.

DiGiovanni GD, Hashemi FH, Shaw NJ, Abrams FA, LeChavallier MW, Abbaszadegan M. Detection of infectious *Cryptosporidium parvum* oocysts in surface and filter backwash samples by immunomagnetic separation and integrated cell culture-PCR. Appl Environ Microbiol 1999;65(8):3427-32.

Dombek, P.E., L.K. Johnson, S.T. Zimmerley, and M.J. Sadowsky. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl Environ Microbiol 2000;66(6): 2572-7.

Dupray E, Caprais MP, Derrien A, Fach P. *Salmonella* DNA persistence in natural seawaters using PCR analysis. Appl Microbiol 1997;82(4): 507-10.

Edwards DD, McFeters GA, Venkatesan MI. Distribution of Clostridium perfringens and fecal sterols in a benthic coastal marine environment influenced by the sewage outfall from McMurdo Station, Antarctica. Appl Environ Microbiol 1998;64(7): 2596-600.

Evangelopoulos A, Legakis N, Vakalis N. Microscopy, PCR and ELISA applied to the epidemiology of amoebiasis in Greece. Parasitol Int 2001;50(3):185-9.

Feria C, Machado J, Duarte Correia J, Goncalves J, Gaastra W. Distribution of papG alleles among uropathogenic *Escherichia coli* isolated from different species. FEMS Microbiol Lett. 2001;202(2): 205-8.

Fournier S, Liguory O, Santillana-Hayat M, Guillot E, Sarfati C, Dumoutier N, et al. Detection of microsporidia in surface water: a one-year follow-up study. FEMS Immunol Med Microbiol. 2000;29(2):95-100.

Frankel G, Riley L, Giron JA, Valmassoi J, Friedmann A, Strockbine N, et al. Detection of *Shigella* in feces using DNA amplification. J Infect Dis 1990;161(6):1252-6.

Franzen C, Muller A. Molecular techniques for detection, species differentiation, and phylogenetic analysis of microsporidia. Clin Microbiol Rev 1999;12(2):243-85.

Furuse KA, Ando A, Osawa S, Watanabe I. Distribution of ribonucleic acid coliphages in raw sewage from treatments in Japan. Appl Environ Microbiol 1981;41:1139-43.

Gardner TB, Hil DR. Treatment of giardiasis. Clin Microbiol Rev 2001;14(1): 114-28.

Geldreich, EE. Sanitary significance of fecal coliform in the environment. U.S. Department of the Interior. Cincinnati, Ohio. Water Pollution Control Research Series Publication, WP-20-3. Federal Water Pollution Control Administration. 1966.

Geldreich EE, Kenner BA. Concepts of fecal streptococci in stream pollution. J Wat Pollut Control Fed 1969;41:R336-R352.

Gerba CP. Phage as indicators of fecal pollution. In *Phage Ecology*, Editors. S.M. Goyal, C.P. Gerba, and G. Bitton. Wiley-Interscience, New York. Pp. 197-209. 1987.

Gratacap-Cavallier B, Genoulaz O, Brengel-Pesce K, Soule H, Bost M, Gofti L, et al. Detection of human and animal rotavirus sequences in drinking water. Appl Environ Microbiol 2000;66(6):2690-2.

Gu W, Aguirre GD, Ray K. Detection of single-nucleotide polymorphism. Biotechniques 1998;24(5):836-7.

Hansson L, Wallbrandt P, Andersson JO, Bystrom M, Backman A, Carlstein A, et al. Carbohydrate specificity of the *Escherichia coli* P-pilus papG protein is mediated by its N-terminal part. Biochem Biophys Acta 1995;1244(2-3):377-83.

Harris SL, Spears PA, Havell EA, Hamrick TS, Horton JR, Orndoff PE. Characterization of *Escherichia coli* type 1 pilus mutants with altered binding specificities. J Bacteriol 2001;183(13):4099-102.

Harwood VJ, Whitlock J, Withington V. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters. Appl Environ Microbiol 2000;66(9):3698-704.

Haslam DB, Boren T, Falk P, Ilver D, Chou A, Xu Z, et al. The amino-terminal domain of the P-pilus adhesin determines receptor specificity. Mol Microbiol 1994;14(3):399-409.

Holcomb DL, Smith MA, Ware GO, Hung YC, Brackett RE, Doyle MP. Comparison of six dose-response models for use with food-borne pathogens. Risk Anal 1999;19(6):1091-1100.

Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci 1981;78(6):3824-28.

Hsu SC, Tsen HY. PCR primers designed from malic acid dehydrogenase gene and their use for detection of *Escherichia coli*. Int J Food Microbiol 2001;64(1-2):1-11.

Johnson JR, Brown JJ. A novel multiply primed polymerase chain reaction assay for identification of variant papG genes encoding the Gal(alpha 1-4)Gal-binding PapG adhesins of *Escherichia coli*. J Infect Dis. 1996;173(4):920-6.

Johnson, JR, Stell, AL, Delavari, P. Canine feces as a reservoir of extraintestinal pathogenic *Escherichia coli*. Infect Immun 2001; 69(3): 1306-14.

Katyal R, Rana SV, Singh K. Rotavirus infections. Acta Virol 2000; 44(5):283-8.

Kingsley DH, Richards GP. Rapid and efficient extraction method for reverse transcription-pcr detection of hepatitis A and norwalk-like viruses in shellfish. Appl Environ Microbiol 2001;67(9):4152- 4157.

Kiss A, Sain B, Venetianer P. The number of rRNA genes in *Escherichia coli*. FEBS Lett 1977;79:77-79.

Kittigul L, Raengsakulrach B, Siritantikorn S, Kanyok R, Utrarachkij F, Diraphat P, et al. Detection of poliovirus, hepatitis A virus and rotavirus from sewage and water samples. Southeast Asian J Trop Med Public Health 2000;31(1):41-6.

Klemm, P. The *FimA* gene encoding the type-1 fimbrial subunit of *Escherichia coli*. Eur J Biochem 1984;143:395-99.

Knolle H. Transmission of poliomyelitis by drinking water and the problem of prevention. Gesundheitswesen 1995;57(6):351-4.

Kostman JR, Edlind TD, LiPuma JJ, Stull TL. 1992. Molecular epidemiology of *Pseudomonas cepacia* determined by polymerase chain reaction ribotyping. J Clin Microbiol 30(8):2084-87.

Latham RH, Stamm WE. Role of fimbriated *Escherichia coli* in urinary tract infections in adult women: correlation with localization studies. J Infect Dis 1984;149:835-40.

Legeay O, Caudrelier Y, Cordevant L, Rigottier-Gois, Lange M. Simplified procedure for detection of enteric pathogenic viruses in shellfish by RT-PCR. J Virol Methods 2000;90(1):1-4.

Le Guyader, F, Haugarreau L, Miossec L, Duboisand E, Pommepuy M. Three-year study to assess human enteric viruses in shellfish. Appl Environ Microbiol 2000;66(8):3241-8.

Li B, Koch WH, Cebula TA. Detection and characterization of the fimA gene of *Escherichia coli* O157:H7. Mol Cell Probes. 1997;11(6):397-406.

Lowery CJ, Moore J, Miller B, Burke D, McCorry K, Crothers E, et al. Detection and speciation of *Cryptosporidium spp.* in environmental water samples by immunomagnetic separation, PCR and endonuclease restriction. J Med Microbiol 49(9): 779-85.

Marc, D. Dho-Moulin M. Analysis of the fim cluster of an avian O2 strain of *Escherichia coli*: Serogroup-specific sites wihin fimA and nucleotide sequence of fimI. J Med Microbiol 1996;44:444-52.

Marklund BI, Tennent JM, Garcia E, Hamers A, Baga M, Lindberg F, et al. Horizontal gene transfer of the *Escherichia coli* pap and prs pili operons as a mechanism for the development of tissue-specific adhesive properties. Mol Microbiol 1992;6(16):2225-42.

Marshall M, Naumovitz D, Ortega Y, Sterling C. Waterborne Protozoan Pathogens. Clinical Microbiology Reviews. 1997;10(1):67-85.

Matar MM, Swaminathan B, Hunter SB, Slater LN, Welch DF. Polymerase chain reaction-based restriction fragment length polymorphism analysis of a fragment of the ribosomal operon from *Rochalimaea* species for subtyping. 1993;31(7): 1730-1734.

Moura H, Sodre FC, Bornay-Llinares FJ, Leitch GJ, Navin T, Wahlquist, S, et al. Detection of an immunofluorescence test of Encephalitozoon intestinalis spores in routinely formalin-fixed stool samples stored at room temperature. J. Clin. Microbiol. 1999;37(7):2317-22.

Murray P, Rosenthal KS, Kobayashi GS, Pfaller MA. Medical Microbiology. Mosby-Year Book, Inc., St. Louis. 2001.

Natarajan A, Souza RE, Lahoti NG, Candrakala SR. Ruptured liver abscess with fulminant amoebic colitis: case report with review. Trop Gastroenterol 2000;21(4):201-3.

Nicholls L, Grant TH, Robins-Browne R. Identification of a novel genetic locus that is required for in vitro adhesion of a clinical isolate of enterohaemorrhagic *Escherichia coli* to epithelial cells. Mol Microbiology 2000;35(2):275-88.

Olsen, JE, Brown DJ, Baggesen DL, Bisgaard M. Biochemical and molecular characterization of *Salmonella enterica* serovar berta, and comparison of methods for typing. Epidemiol Infect 1992;108: 243-60.

Olsen PB, Schembri MA, Gally DL, Klemm P. Differential temperature modulation by H-NS of the fimB and fimE recombinase genes which control the orientation of the type 1 fimbrial phase switch. FEMS Microbiol Lett 1998;162(1):17-23.

Orndorff PE, Falkow S. Nucleotide sequence of pilA, the gene encoding the structural component of type 1 pili in *Escherichia coli*. J Bacteriol 1985;162(1):454-57.

Orskov F, Orskov I. In: A. I. Broude, editor. Enterobacteriaceae. Medical microbiology and infectious diseases. Philadelphia: W.B. Saunders Corp;1981.

Orskov I, Orskov F. Serologic classification of fimbriae. Curr Top Microbiol Immunol 1990;151:71-90.

Parveen S, Hodge NC, Stall RE, Farrah SR, Tamplin ML. Phenotypic and genotypic characterization of human and nonhuman *Escherichia coli*. Water Res. 2001;35(2):379-86.

Parveen S, Murphree R, Edminston L, Kaspar CW, Portier KM, Tamplin ML. Association of multiple-antibiotic resistance profiles with point and non point sources of *Escherichia coli* in Apalachicola bay. Appl Environ Microbiol 1997;63:2607-12.

Parveen S, Portier KM, Robinson K, Edminston L, Tamplin ML. Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Appl Environ Microbiol 1999;65:3142-47.

Peek AS, Souza V, Equiarte LE, Gaut BS. The interaction of protein structure, selection, and recombination on the evolution of the type-1 fimbrial major subunit (fimA) from *Escherichia coli*. J Mol Evol 2001;52:193-204.

Pina S, Puig M, Lucena F, Jofre J, Girones R. Viral pollution in the environment and in shellfish: human adenovirus detection by PCR as an index of human viruses. Appl Environ Microbiol 1998;64(9):3376-82.

Popovic T, Bopp C, Olsvik O, Wachsmuth K. Epidemiologic application of a standardized ribotype scheme for *V. cholerae* 01. J Clin Microbiol 1993;31: 2474-82.

Pourcher AM, Devriese LA, Hernandez JF, Delattre JM. Enumeration by a miniaturized method of *E. coli*, *Streptococcus bovis* and enterococci as indicators of the origin of faecal pollution of waters. J Appl Bacteriol 1991;70:525-30.

Puig M, Jofre J, Lucena F, Allard A, Wadell G, Girones R. Detection of adenovirus and enterovirus in polluted waters by nested PCR amplification. Appl Environ Microbiol 1994;60(8):2963-70.

Reynolds KA, Gerba CP, Abbaszadegan M, Pepper LL. ICC/PCR detection of enteroviruses and hepatitis A virus in environmental samples. Can J Microbiol 2001;47(2):153-7.

Rochelle PA, De Leon R, Johnson A, Stewart MH, Wolfe RL. Evaluation of immunomagnetic separation for recovery of infectious *Cryptosporidium parvum* oocysts from environmental samples. Appl Environ Microbiol 1999;65(2):841-5.

Rochelle PA, Ferguson DM, Handojo TJ, De Leon R, Stewart MH, Wolfe RL. An assay combining cell culture with reverse transcriptase PCR to detect and determine the infectivity of waterborne Cryptosporidium parvum. Appl Environ Microbiol 1997;63(5): 2029-37.

Rochelle PA, Ferguson DM, Handojo TJ, DeLeon R, Stewart MH, Wolfe RL. Development of a rapid detection procedure for *Cryptosporidium*, using in vitro cell culture combined with PCR. J Eukaryot Microbiol 1996;43(5):72S.

Roe A, Currie C, Smith D, Gally D. Analysis of type 1 fimbriae expression in verotoxigenic *Escherichia coli*: a comparison between serotypes O157 and O26. Microbiology. 2001;147(1):145-52.

Rose M, Scholler G, Jorres A, Danzer G, Klapp BF. Patients' expressions of complaints as a predictor of the course of acute hepatitis A. J Psychosom Res 2000;48(2):107-13.

Rotbart HA, McCracken GH, Whitley RJ, Modlin JF, Cascino M, Shah S, et al. Clinical significance of enteroviruses in serious summer febrile illnesses of children. Pediatr Infect Dis 1999;18(10):869-74.

Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual, 2nd edition, Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y. 1989

Savage DC, Fletcher DC. Editors. Bacterial Adhesion: Mechanisms and Physiological Significance. Plenum Press, New York. 1985.

Schembri MA, Kjaergaard K, Sokurenko EV, Klemm P. Molecular characterization of the *Escherichia coli FimH* adhesin. J Infect Dis 2001;183:S28-S31.

Schembri MA, Olsen PB, Klemm P. Orientation-dependent enhancement by H-NS of the activity of the type 1 fimbrial phase switch promoter in *Escherichia coli*. Mol Gen Genet 1998;259(3): 336-344.

Schwab KJ, Neill FH, Le Guyader FF, Estes MK, Atmar RL. Development of a reverse transcription-pcr-dna enzyme immunoassay for detection of "norwalk-like" viruses and hepatitis a virus in stool and shellfish. Appl Environ Microbiol 2001;67(2):742-9.

Schunk M, Jelinek T, Wetzel K, Nothdurft HD. Detection of *Giardia lamblia* and *Entamoeba histolytica* in stool samples by two enzyme immunoassays. Eur J Clin Microbiol Infect Dis 2001;20(6);389-391.

Seas C, Miranda C, Gil AI, Leon-Barua R, Patz J, Huq A, et al. New insights in the emergence of cholera in Latin America during 1991: the Peruvian experience. Am J Trop Med Hyg 2000;62(4):513-570.

Sekizaki T, Ito H, Asawa T, Nonomura I. DNA sequence of type 1 fimbrin, Fpul1, gene from a chicken pathogenic *Escherichia coli* seortype O78. J Vet Med Sci 1993;55(3):395-400.

Stull TL, LiPuma JJ, Edlind TD. A broad-spectrum probe for molecular epidemioilogy of bacteria; ribosomal RNA. J Infect Dis 1988;157: 280- 86.

Stutman HR. *Salmonella, Shigella*, and *Campylobacter*: common bacterial causes of infectious diarrhea. Pediatr Ann 1994;23(10): 538-43.

Svec P, Sedlacek I, Pantucek R, Devriese LA, Doskar JV. Evaluation of ribotyping for characterization and identification of Enterococcus haemoperoxidus and Enterococcus moraviensis strains. FEMS Microbiol Lett. 2001;203(1):23-7.

Szakal D, Gado I, Pat T. A colony blot immunoassay to detect enteroinvasive *Escherichia coli* and *Shigella* in water samples. Appl. Microbiol. 2001;90(2):229-36.

Tamplin ML, Jackson JK, Buchreiser C, Murphree RL, Portier KM, Ganger VV, et al. Pulsed-field gel electrophoresis and ribotype profiles of clinical and environmental *Vibrio vulnificus* isolates. Appl Environ Microbiol 1996;62:3572-80.

Tewari R, MacGregor JI, Ikeda T, Little JR, Hultgren SJ, Abraham SN. Neutrophil activation by nascent fimH subunits of type 1 fimbriae purified from the periplasm of *Escherichia coli*. J Biol Chem 1993; 268:3009-15.

Tullus K, Kuhn I, Orskov I, Orskov F, Mollby R. The importance of P and type 1 fimbriae for the persistence of *Escherichia coli* in the human gut. Epidemiol Infect 1992;108(3):415-21.

USEPA. ICR protozoan method for detecting Giardia cysts and Cryptosporidium oocysts in water by a fluorescent antibody procedure. Office of Ground Water and Drinking Water. Washington, D.C. 1995.

USEPA. Method 1623: Cryptosporidium and Giardia in water by filtration/IMS/FA. Office of Water. Washington, D.C. 1999.

Versalovic J, de Bruijn FJ, Lupski JR. Repetitive sequence-based PCR (rep-PCR) DNA fingerprinting of bacterial genomes. In: F.J. de Bruijn, J.R. Lupski and G.M. Weinstock, editors. Bacterial genomes: physical structure and analysis. New York: Chapman and Hall, 1998. p. 437-454.

Versalovic J, Schneider M, DeBruijn FJ, Lupski JR. Genomic fingerprinting of bacteria using repetitive sequence based polymerase chain reaction. Methods Mol Cell Biol 1994;5:25-40.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995;23(21):4407-14.

Waage AS, Vardund T, Lund V, Kapperud G. Detection of small numbers of *Campylobacter jejuni* and *Campylobacter coli* in environmental water, sewage, and food samples by semi-nested PCR assay. Appl Environ Microbiol 1999;65(4):1636-43.

Wachsmuth K. Molecular epidemiology of bacterial infections: example of methodology and investigations of outbreaks. Rev Infect Dis 1986;8:682-92.

Watanabe G, Umetsu K, Yuasa I, Sato M, Sakabe M, Naito E et al. A novel technique for detecting single nucleotide polymorphisms by analyzing consumed allele-specific primers. Electorphoresis 2001;22 (3): 418-20.

Wiggins BA. Discriminant analysis of antibiotic resistance patterns of fecal Streptococci: a method to differentiate human and animal sources of fecal pollution in natural waters. Appl Environ Microbiol 1996;62:3997-4002.

Wiggins BA, Andrews RW, Conway RA, Corr CL, Dobratz EJ, Dougherty DP, et al. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Appl Environ. Microbiol. 1999; 65:3484-6.

Yu H, Bruno JG. Immunomagnetic-electrochemiluminescent detection of *Escherichia coli* 0157:H7 and *Salmonella typhimurium* in foods and environmental water samples. Appl Environ Microbiol 1996; 62(2): 587-92.

Yuri K, Nakata K, Katae H, Yamamoto S, Hasegawa A. Distribution of uropathogenic virulence factors among *Escherichia coli* strains isolated from dogs and cats. J Vet Med Sci 1998; 60(3): 387-90.

Zierdt CH, Robertson EA, Williams RL, Maclowry JD. Computer analysis of Staphylococcus aureus phage typing data from 1957-1975, citing epidemiological trends and natural evolution within phage typing system. Appl Environ Microbiol 1980;39:623-29.

Zindrou S, Orozco E, Linder E, Tellez A, Bjorkman A. Specific detection of *Entamoeba histolytica* DNA by hemolysin gene targeted PCR. Acta Trop 2001;78(2);117-25.

BIOGRAPHICAL SKETCH

Troy M. Scott was born in Toledo, Ohio on March 20, 1975. In 1993, he graduated with honors from Wellington High School in Wellington, Florida. After high school, he attended the University of Florida and graduated with honors with a Bachelor of Science in 1997 and a Master of Science in 2000. After graduation, he will continue his education with postdoctoral studies in Environmental Microbiology.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Samuel R. Farrah
Samuel R. Farrah, Chair
Professor of Microbiology
and Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Phillip Achey
Phillip Achey
Professor of Microbiology
and Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

E M Hoffmann
Edward M. Hoffmann
Professor of Microbiology
and Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Paul A. Gulig
Professor of Molecular
Genetics and Microbiology

This dissertation was submitted to the Graduate Faculty of the College of Agricultural and Life Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy

May 2002

Dean, College of Agricultural
and Life Sciences

Dean, Graduate School

LD
1780
2002
.S429

UNIVERSITY OF FLORIDA
3 1282 08556 8148

LD
1780
2002
.S429

UNIVERSITY OF FLORIDA
3 1282 08556 8148